

Time to cut regulations that protect only regulators

Researchers and their live subjects in US universities are valuably protected by some regulatory processes but pointlessly undermined by others. It is time to streamline or scrap the latter.

The tragic deaths of two young people in clinical trials at the University of Pennsylvania and Johns Hopkins University have had a chilling impact on the scientific community, by highlighting what can happen when well-intentioned research goes wrong. Attention in the United States has correspondingly been focused on the workings of the Institutional Review Boards (IRBs) that review and approve human experimentation.

No doubt all this will mean even more federally mandated rules and regulations that US scientists and institutions will have to follow in order to conduct scientific research legally. Unfortunately, what seems to happen whenever such controversies arise is a blanket application of 'the rules', with no leeway for common sense. The overall effect is that researchers and institutions alike are drowning in paperwork. Sadly, much of the documentation bears little relation to the realities of research, for it accomplishes little beyond slowing down legitimate experiments and making them harder to administer. The work still goes on — just more slowly, and at greater cost in time and money.

Such views will be discounted by some as reflecting the blinkered interests of researchers. *Nature* has not hesitated in the past to urge scientists to be more responsive to ethical and social concerns in this context. But it is important to discriminate between what is truly a risk to the study subject, whether human or animal, and what is not. The former category of experiment requires rigorous review, to prevent unnecessary deaths and injury. But much of the research that falls under the purview of IRBs and Institutional Animal Care and Use Committees (IACUCs) does not warrant such close scrutiny, as it causes little real risk for the study subject or, in the case of animals, involves generally accepted procedures such as immunization, taking blood or challenge with standard pathogens. The time spent by committees reviewing such routine protocols to the full extent required by law is time that could be better spent on the more controversial applications.

Climate of fear

This point has been well made in the context of IRB reviews by Robert Schooley, of the University of Colorado School of Medicine, and colleagues in various publications, but it applies to animal research as well. In one respect, the situation with IACUCs is even worse than with IRBs. Nowadays, IACUC reviews must take place before the grant application can even be peer reviewed by the National Institutes of Health (NIH). As most applications are rejected upon peer review, an enormous amount of time is wasted by IACUCs in approving animal experiments that will never be performed. The procedure used to be that an IACUC application was only necessary once a grant application had fared well enough to be approved in principle; this sensible process was changed through pressure on the NIH from animal-welfare lobbyists, leading to yet more paperwork but no real benefit to animals.

The climate under which even routine protocols are reviewed by IACUCs and IRBs is now one of fear — fear by the institution that it

will be "out of compliance" with one or more aspects of the paperwork, and so subject to penalty upon audit (be that by the NIH, the Office for Human Research Protection, the US Department of Agriculture, or whatever other organization is involved). Again, the attention is on the paperwork, not on the truly important question: is the risk of the experiment to the subject worth the likely benefit of the experiment to society?

Paper-trail politics

In the legal system, precedents are important criteria for determining the outcome of cases, but this does not apply with review committees: even if a similar experiment has been approved and successfully carried out elsewhere it must be reviewed yet again in-house. This is not, in reality, to protect the animal — it is to protect the institution and, by extension, the reviewers.

There is no permitted cross-talk between committees in different institutions and no central list of approved procedures, so each institutional committee must act as an island unto itself. Indeed, if an experiment involves a collaboration between institutions, each institution (perhaps as many as three or four) must individually review the application — adding to the cost of the review but not to the welfare of the humans or animals involved. Non-controversial applications are multiply nit-picked; forms are returned to the applicant because a single word has been mistyped, or even because the wrong font has been used.

It seems as if review committees 'must' find something wrong with each and every application, so as to generate a paper trail proving the thoroughness of their review. For controversial applications this is entirely appropriate, but the 'fear factor' means that enormous effort is now expended on routine and mundane protocols. In some institutions, scientists who never work with organisms more complex than yeast and bacteria are now being forced to attend lectures on how to conduct research on humans, "just in case". This is yet another institutional response to pressure from federal agencies, but what purpose does it serve in the real world?

Does society want to support scientific research? If it does, it should find ways to make the system work efficiently as well as safely. At present, local and federal administrators and scientists alike are frustrated by what must be done to make research legal in the face of those who would have it otherwise.

There's also a price to pay for the extra workload — committee member burn-out. To use a topical analogy, how many suitcases full of underwear and socks can an airline security agent wade through before he or she is too tired to find the hidden box-cutter? It's more efficient, surely, to identify the threatening passenger before the suitcase is packed, than to pick every case apart, just to check if there's a problem. Just as travellers have to travel, scientists still have to do science. It's a sad day when the working scientist realizes that the most important person in the laboratory is not the smartest postdoc nor the hardest-working student, but the chief administrator. ■





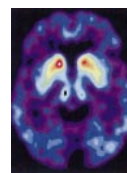
Down to Earth
Europe gets tough
over spending on
space projects
p383



Bleak outlook
Dispute over website
leaves Genome
Database faltering
p384



Sinking feeling
Research projects for
carbon sinks get
under way
p385



Fetal attraction
Boost for
transplants to
treat Parkinson's
p386

Imploding detectors shatter plans for Japan's neutrino experiments

David Cyranoski, Tokyo

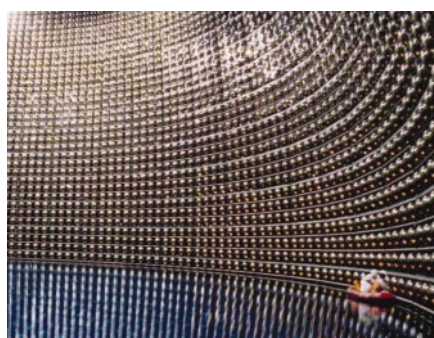
Neutrino physicists around the world have been stunned by an accident at Japan's Super-Kamiokande experiment. The damage will put 'Super-K' out of action for at least a year, delaying several research projects.

Super-K consists of a tank filled with 50,000 tonnes of water, buried deep under Mount Ikeno, some 240 kilometres north-west of Tokyo. Neutrinos from the Sun and from supernovae, and those formed when cosmic rays strike the atmosphere, barely interact with matter. But on the rare occasions that they collide with electrons, such as those in water molecules, they emit flashes of blue light known as Cerenkov radiation. These can be detected by sensors called photomultiplier tubes (PMTs).

The walls of Super-K were lined with 11,200 PMTs. Since August, the tank had been drained for maintenance. But on 12 November, as it was being refilled, many of the PMTs imploded. "This has shaken the neutrino physics community," says Francis Halzen of the University of Wisconsin in Madison.

Researchers familiar with such projects say that the pressure of the water may have burst one of the lower sensors, sending shock waves into neighbouring PMTs and bringing about a chain reaction. About 7,000 of the PMTs shattered, and technicians are now conducting tests to see how many of the remaining PMTs are still functioning. Although the accident also appears to have produced a small crack in the tank's lining, resulting in minor leakage, this is expected to require only minor repairs.

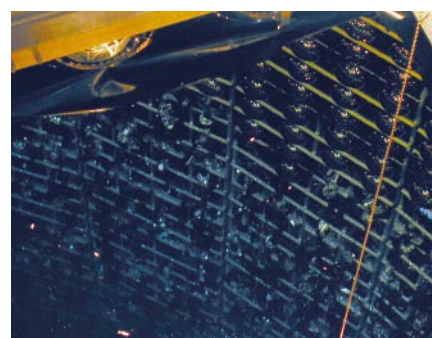
Yoji Totsuka, director of the Kamioka Observatory, which runs Super-K, and head of Tokyo University's Institute for Cosmic Ray Research, is urgently attempting to determine the cause of the disaster. He is also trying to reassure international collaborators. "We will rebuild the detector. There is no question," reads a message on Super-K's website. The education ministry, which oversees the facility, has promised support, but says it will not ask for special funds. This means that the costs of the repairs — estimated at between US\$10 million and US\$30 million — will have to be



Gone in a flash: Super-K (left) before the accident that left many photomultiplier tubes in pieces (right).

taken from other projects.

The PMTs used at Super-K were made by Hamamatsu Photonics. At more than 50 centimetres across, they are twice the diameter of PMTs at most other neutrino facilities. Totsuka acknowledges that the sensors' large size may have made them more fragile. But he



argues that the choice was justified by their improved sensitivity. "The advantage is obvious if you consider the results coming out of the experiment," Totsuka told *Nature*.

In 1998, Super-K produced the long-awaited, ground-breaking evidence suggesting that neutrinos have mass. But its work is

Loophole legalizes human cloning

David Adam, London

Britain's cloning regulations have been thrown into chaos by a court ruling that it is not illegal to create a cloned human baby. In a judgement on 15 November, the High Court ruled that existing laws do not cover cloned embryos because they are not formed through fertilization. The government says it will appeal, and is also considering emergency legislation.

Earlier this year the 1990 Human Fertilisation and Embryology Act was amended (see *Nature* 409, 5; 2001) to allow research into using human embryonic stem (ES) cells to grow replacement tissues. This was intended to include cloned embryos, to allow work on 'therapeutic' cloning, which holds the prospect of growing grafts perfectly matched to individual patients. No one has yet applied to the Human Fertilisation and Embryology Authority (HFEA) for a licence to conduct such research, but the law has been praised by stem-cell researchers for appreciating the

difference between therapeutic and reproductive cloning.

The new ruling — on a case brought by the ProLife Alliance, which opposes ES-cell research — means that the HFEA no longer has the jurisdiction to award such licences. It effectively leaves human cloning completely unregulated. Severino Antinori, the Italian fertility expert who claims that he will clone human babies, says he now wants to relocate to Britain.

The court's decision will not have an immediate impact on British ES-cell researchers, as they are currently working with cells isolated from embryos created by *in vitro* fertilization. But they say it is important that the cloning regulations are restored. Austin Smith, director of the Centre for Genome Research at the University of Edinburgh, said in a statement that if the court has "identified a loophole that would permit the creation of children by reproductive cloning, it is essential that such a loophole be blocked".

not finished, and the accident will hamper continuing projects. "We can only hope that they are able to restore the detector soon," says Art McDonald, director of the Sudbury Neutrino Observatory in Ontario, Canada.

The project that will feel the most immediate impact is 'K2K', which shoots a beam of neutrinos 250 kilometres through the Earth's crust from an accelerator at the High Energy Accelerator Research Organization (KEK) in Tsukuba, north of Tokyo. This Japan-US-Korea collaboration aims to study neutrino 'oscillation', the conversion of neutrinos from one form to another which confirms they have mass.

Neutrinos come in three forms, or 'flavours', and in experimental runs up to April this year, muon neutrinos, the flavour shot from KEK, have been detected by Super-K's sensors 44 times. As this falls short of the 66 expected sightings, the results suggest that the particles are turning into another neutrino variety. "This is an indication, but not conclusive evidence," admits Koichiro Nishikawa, a spokesman for the K2K collaboration.

The plan was to triple the amount of data by 2005, when the proton synchrotron used in the experiment will be shut down. K2K was scheduled to start taking data again in January 2002. Totsuka hopes to have Super-K back in action by January 2003. If so, it may be possible to increase data collection from the originally planned six months per year to eight or more months to make up for the lost time.

Totsuka is now trying to find a quick and relatively cheap way to get Super-K back online. The damaged PMTs cost \$3,000 each when they were installed during 1992-96, and Hamamatsu discontinued their production in 1998. One possibility is to halve the number of PMTs. This would reduce Super-K's sensitivity to solar neutrinos and, to a lesser extent, to atmospheric neutrinos. But those released from KEK are of sufficiently high energy for them still to be readily detected.

Totsuka also hopes that the cheaper, more advanced PMTs being developed for a proposed upgrade known as Hyper-Kamiokande, planned to start within a decade or more, might be introduced earlier at Super-K.

The world's neutrino physicists are hoping that the damage to Super-K can soon be repaired. Its absence from the field will be keenly felt, they say. "I am emotionally attached to the Super-K detector," says John Bahcall of the Institute for Advanced Study in Princeton, New Jersey. "It has done things that I only previously dreamed about."

www-sk.icrr.u-tokyo.ac.jp

Genetics paper erased from journal over political content



War of words: controversial descriptions of the Israeli-Palestinian conflict led to the article being axed.

Erica Klarreich

A paper about the genetic origins of Palestinians has found itself at the centre of a political storm. In a highly unusual move, the journal *Human Immunology* has deleted the paper from its September issue after receiving a wealth of complaints over what some saw as inappropriate political comments about the Israeli-Palestinian conflict.

The paper examines genetic variability in the HLA complex — a highly diverse complex of immune-system genes — in a sample of Palestinians (A. Arnaiz-Villena *et al. Hum. Immunol.* **62**, 889-900; 2001). But controversially, it also includes a historical introduction calling Jews living in the Gaza strip "colonists" and describing some Palestinians as living in concentration camps. The paper's publication sparked a "cascade" of angry letters complaining that such comments had no place in a scientific journal, says the journal's editor-in-chief, Nicole Suciu-Foca of Columbia University in New York.

The paper "purports to be a scientific treatise" but "offers opinion on geopolitical issues that cannot be substantiated by the data presented", wrote Dolly Tyan, then president of the American Society for Histocompatibility and Immunogenetics (ASHI), which runs the journal, in a letter to members on 3 October. "ASHI is offended and embarrassed by its inclusion within the journal."

The publisher of *Human Immunology*, Elsevier Science, has removed all electronic versions of the article and has sent a letter to individual subscribers and librarians advising them to ignore the article "or, preferably, to physically remove the relevant pages".

The paper's lead author, Antonio Arnaiz-Villena of the Complutense University in Madrid, says he did not intend to offend anyone and calls the decision to withdraw the article "unwise". He says he has several letters of support, including one from Jean Dausset,

president of the Human Polymorphisms Study Centre in Paris, one of the founding fathers of HLA genetics. A more appropriate action, Arnaiz-Villena says, would have been to publish the letters of complaint and allow him to respond.

But the depth of anger the article raised made such a course impossible, argues Suciu-Foca. One ASHI member was so offended by the article that he resigned, she says. "We would have had mass resignations and the journal would have been destroyed if this paper were allowed to remain."

The paper was in a special issue on anthropology edited by Arnaiz-Villena. Although Arnaiz-Villena says the paper was approved by two reviewers, the incident has prompted the journal's editorial board to revise its policy so that in future the editor-in-chief will supervise work by guest editors, Suciu-Foca says.

The data announced in the paper, which indicated that Jews and Palestinians have a close genetic relationship, were worth reporting, says Steven Marsh, a member of *Human Immunology*'s editorial board who studies the nomenclature of HLA genes at the Anthony Nolan Research Institute in London. "Had the authors confined themselves to announcing their scientific results, it would have been an interesting paper," he says.

The retraction of a scientific paper because of political statements is "unprecedented", says Sheldon Krimsky, an expert on publication ethics at Tufts University in Medford, Massachusetts. But the editorial board took legal advice before making its decision, Suciu-Foca says. "This has nothing to do with freedom of opinion," she says. "This journal is not the right forum for expressing political views."

The editors invited Arnaiz-Villena to resubmit a revised version of the article, Suciu-Foca says, and reviewers are now considering a new version for possible publication.

Europe puts the squeeze on space projects

Sally Goodman

Europe's space scientists are facing some tough decisions, after the member states of the European Space Agency (ESA) agreed a five-year budget that will require some planned missions to be cut.

Meeting in Edinburgh last week, ministers from 15 European nations and Canada agreed funding for ESA's entire range of projects. They committed 1.869 billion euros (US\$1.6 billion) over five years to the agency's science programme, into which all members pay a mandatory subscription according to their wealth. This maintains roughly level funding, including a modest compensation for inflation. But David Southwood, director of ESA's science programme, had asked for 1.945 billion euros. "The outcome is satisfactory, but a big loss on what we were hoping for," he says.

Southwood speculates that a single big project is most likely to be sacrificed. He says the GAIA satellite, which would map the distances and positions of the billion brightest objects in the sky and is scheduled for launch by 2012, "is in some jeopardy". Lennart Lindgren of the Lund Observatory in Sweden, one of GAIA's principal investigators, says this would be "extremely disappointing".

Preparation for the Aurora programme, an ambitious long-term plan for robotic and human exploration of the Solar System (see *Nature* **411**, 625; 2001), was one of the non-mandatory programmes given the go-ahead, but in reduced form. Delegates were shocked when Italy, which had been Aurora's main



Goodbye? Edelgard Bulmahn with ISS crew (from left) Sergei Krikalev, Yuri Gidzenko, Bill Shepherd.

advocate, came to the table empty-handed.

Italy had originally offered to pay 40% of the 40 million euros needed for a three-year preparation programme. But given the current squeeze on research spending (see page 384), "Italy must be content to be one of the participants in a very reduced global programme", says Giovanni Bignami, scientific director of the Italian space agency. Under pressure from other ESA members, the Italian delegation finally agreed to contribute 2 million euros towards a total of 14 million euros. France, Britain, Spain, Belgium, Switzerland, Austria, Portugal and Canada also subscribed to the programme.

In a strong message to the United States, the ministers decided to hold back 60% of ESA's exploitation budget for the International Space Station (ISS), pending NASA's

agreement to fulfil its original agreement to operate the ISS with a full crew of six. Without this, it will be hard to carry out scientific research. "Europe will fulfil its obligations and expects the same from our partners," said German science minister Edelgard Bulmahn, who chaired the meeting.

But given the nomination of a NASA administrator who appears determined to cap the space station's budget (see below), it is unclear whether ESA's move will have the desired effect. And there were signs that ESA members have doubts about the station's scientific potential. The ELIPS programme, which will support experiments in both physical and life sciences on board the ISS, received only 166.5 million euros of the requested budget of 320 million euros. ■

♦ www.esa.int

Budgets come first as Bush chooses NASA head

Mark Schroppe

President George W. Bush has sent NASA a message about costs by nominating a budget expert with no aerospace or scientific background as the agency's administrator.



The nominee, Sean O'Keefe, a former secretary of the Navy and chief financial officer with the Department of Defense, was appointed as deputy director at the Office of Management and Budget in March. There he scrutinized overspending in the International Space Station programme, which he termed a "management and financial crisis".

If his appointment is confirmed, as expected, by the Senate, O'Keefe will have to implement reforms he recommended in his previous position, making working within budgets as important as technical excellence.

John Pike, a former head of space policy with the Federation of American Scientists, says that the Bush administration's policy is centred on the need to control the space station's budget. "The space programme has not been a priority for them," he says. "It was on the backburner before September and it isn't even on the stove now."

"It's an interesting choice," observes John

Logsdon, director of the Space Policy Institute at George Washington University in Washington, DC. "They are taking someone out of the centre of the Bush administration and sending him over to NASA to fix some perceived problems."

Scientists are not sure what to make of the nomination. "I was quite startled to begin with," says Anneila Sargent of the California Institute of Technology in Pasadena, president of the American Astronomical Society. "But on reflection I have come to see that it might have very positive aspects." Because space science has traditionally managed its finances well, she says, it may find favour with a budget-conscious chief.

But Sargent is concerned that budget issues could override scientific and technological arguments altogether. "I think there has to be a balance," she says. "We can't do everything we want, but we can't be told that we can't do anything." ■

Sackings leave gene database floundering

Laura Bonetta

Just a decade ago, it was at the cutting edge of bioinformatics. But today, the Genome Database (GDB) — which is supposed to be the main international repository for genome mapping information — cuts a sad figure.

The Hospital for Sick Children (HSC) in Toronto, Canada, which hosts the database, has fired the GDB's director, Jamie Cuticchia, and another official, Gregg Silk, alleging that they illegally transferred the GDB's Internet domain name to a Maryland-based non-profit company. Cuticchia argues that the transfer was part of a legitimate effort to raise funds.

The GDB was established at Johns Hopkins University in Baltimore, Maryland, in 1990, with funding from government agencies including the US Department of Energy and the National Institutes of Health. The GDB's mission, according to its website, is "to make available to scientists an encyclo-

pedia of the human genome that is being constantly revised and updated".

Cuticchia moved the database from Johns Hopkins to Toronto in 1998 after the US Department of Energy, then the primary funder, cut off support for the project. Cuticchia has raised about Can\$700,000 (US\$441,000) a year in funding to run the GDB, but the database has struggled to maintain itself as a useful resource. It has not incorporated sequence data to supplement its maps, for instance, and its software tools need upgrading.

"GDB was trying to change its approach and get up to date," says Ewan Birney of the European Bioinformatics Institute in Hinxton, near Cambridge.

The latest moves have thrown the GDB's future into question. Cuticchia maintains that he discussed with HSC officials his plans to redevelop the GDB by selling licences to companies allowing them to run a copy of the database. Cuticchia says



A row has erupted over the Genome Database at the Hospital for Sick Children in Toronto.

Funding fears spark Italian protests

Alison Abbott

One hundred and fifty scientists from Milan took part last week in a symbolic funeral — complete with coffin, musicians and black balloons — for Italian research.

The mock funeral was part of a wave of protests, including strikes and demonstrations, that took place across the country, as scientists expressed their fear that Prime Minister Silvio Berlusconi's right-wing government will reduce support for basic research.

Since taking office in July, Berlusconi's government has unsettled scientists by hinting that it favours applied research. Now rumours that budgets will be cut are being confirmed. And in an unprecedented move, the government last week rejected the three-

year plan of the National Research Council (CNR), which had proposed an expansion of staff. Scientists fear that reforms introduced by the previous government to make Italian research more efficient are now under serious threat (see *Nature* 412, 264–265; 2001).

When public expenditure plans for 2002 are approved towards the end of the year, staffing is expected to be frozen and budget cuts are in store. Sources in the education and research ministry say that spending is likely to be kept at 2001 levels, effectively cutting research budgets by the annual rate of inflation, now running at 2.9%.

Scientists who helped prepare the CNR's three-year plan are dispirited. In the past two years, the agency has reorganized its 330 units into 100 larger institutes. To make this system work efficiently, it planned to increase its staff by 1,750, or 30%, and its budget by 90%, over three years. Elisabetta Visalberghi, research director of the CNR Institute for Psychology in Rome, says the expansion is important to ensure that there are replacements for the generation of baby-boomers who are now reaching retirement.

University reforms are also under pressure. Italian universities are changing from a system of five-year undergraduate degrees to three-year bachelor programmes plus optional two-year masters degrees. The changes will involve an additional teaching load, for which the government does not want to provide extra money.

that the hospital decided not to be involved and told him that he was free to seek investors. Cuticchia says that the transfer of the GDB's domain name, to a company called GDB Human Genome Database, was a step in his plan to raise funds to update the database's software and continue providing academics with free access.

On 18 October, Cuticchia says that he and Silk, who was director of planning and operations of the HSC's bioinformatics centre, told hospital management that they had secured funding for the redevelopment of the GDB. The following day, the pair were suspended and on 2 November they were fired.

But HSC spokeswoman Cyndy DeGiusti says that the hospital only found out about Cuticchia's efforts to commercialize the database "by accident". The hospital has filed a 'notice of action' — the first step towards a lawsuit — to recover the domain name. "We caught it early enough so that the actual value of what was taken was not high, but it was proceeding down the road of taking valuable assets," says DeGiusti.

Cuticchia has been barred from his lab while the hospital conducts an audit. HSC geneticist Lap-Chee Tsui, co-discoverer of the cystic fibrosis gene, will for the time being oversee the GDB.

Although his appointment at the University of Toronto — with which the HSC is affiliated — has not been affected, Cuticchia says he is forbidden to contact his students. "The hospital has set a precedent by saying 'We can take your grants, equipment and students,'" claims Cuticchia. "It is a black day for academic freedom."

♦ www.gdb.org



Funeral party: Italian scientists take to the streets to express their fears for basic research.

Health priorities gain patent reprieve for developing countries

Michael Cherry

In an unexpected move, the World Trade Organization (WTO) last week agreed that patent laws would not be used to prevent developing countries from using generic medicines to protect public health.

The trade-related aspects of intellectual property (TRIPS) agreement currently permits countries to license the production of cheaper copies of patented drugs where this is necessary on public-health grounds. But if the WTO had required all its members to recognize patents by 2006, as was originally planned, these rights could have fallen away.

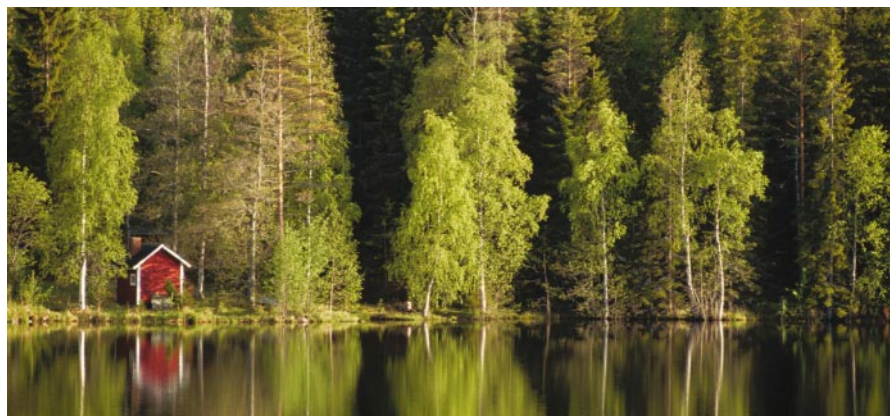
The declaration agreed at the WTO meeting in Doha, Qatar, on 14 November postpones this deadline to 2016 for least-developed countries, and states that the agreement should be implemented in a manner that is "supportive of WTO members' right to protect public health, and in particular to promote access to medicines for all". The agreement was brokered by Brazil — which produces large quantities of generic drugs — and was unexpectedly backed by the United States in the face of opposition from the US pharmaceutical industry.

In addition, individual countries will be free to determine the circumstances under which licences to manufacture generics are issued. The Doha declaration acknowledges that these are not confined to emergency situations — but if countries declare an emergency they can issue compulsory licences without prior negotiation with the patent owner.

The loosely worded TRIPS agreement has been open to different interpretations in the past, leading to threats of trade sanctions and legal actions of the kind that the South African Pharmaceutical Manufacturers Association brought against its government earlier this year for enacting legislation which it claimed threatened intellectual property rights (see *Nature* 410, 1013; 2001). Mark Heywood, national secretary of South Africa's Treatment Action Campaign, hails the new agreement as "a step in the right direction" that should prevent such actions in future.

"Doha sends a strong message that people's health overrides the interests of big drug companies," agrees Michael Bailey, senior policy adviser to the charity Oxfam. But problems will remain for countries that lack a generic drugs industry, and therefore cannot obtain drugs by issuing a compulsory licence. ■

Cycle studies see carbon sinks rise to prominence



Green schemes: two studies will look at how planting forests could reduce atmospheric carbon.

Quirin Schiermeier

As the dust settles on the latest round of negotiations over the Kyoto Protocol on climate change, researchers are turning their attention to one of the most controversial scientific issues behind the accord — the role of sinks such as soils and vegetation in absorbing atmospheric carbon.

Two ambitious studies — one in Europe, the other in the United States — are set to assess how much carbon is absorbed by such sinks. CarboEurope, a cluster of eight related projects, was established last year with the aim of understanding and quantifying the exchange of carbon between vegetation and the atmosphere. A similar programme was proposed last week by the US Carbon Cycle Interagency Working Group, a committee that includes representatives of the major US funding organizations. Both studies are aimed at laying the groundwork for better monitoring networks to verify the protocol once it comes into force.

Carbon sinks are an important component of the Kyoto Protocol. About half of the 6.5 billion tonnes of carbon emitted globally by burning fossil fuels is taken up by vegetation and stored as organic matter. According to the Kyoto rules, the details of which were negotiated earlier this month in Marrakech (see *Nature* 414, 238; 2001), countries receive credits for carrying out activities that increase the absorption of carbon, such as planting forests. The credits can be used to reduce the amount by which they have to cut their emissions, or can be traded on the global emissions market.

But the protocol is vague about how credits for sink activities will be assessed. Governments currently use data from forest inventories, which measure forest areas and timber volume, to determine the effects of sinks. Alternative techniques, such as measuring atmospheric carbon levels, pro-

duce different estimates of sink potential, leaving the process open to manipulation by governments keen to earn credits.

"From a scientific point of view, only a full continental carbon balance can avoid unwanted loopholes," says Ernst-Detlef Schulze, managing director of the Max Planck Institute for Biogeochemistry in Jena, and one of the initiators of CarboEurope.

CarboEurope, which receives 30 million euros (US\$26 million) from the European Commission and national governments, involves around 400 scientists from 75 European institutes. Data from various sources, including soil samples, forest inventories and plant respiration studies, are being made available in consistent formats to cross-check the results of different model calculations of carbon cycles. The programme is expanding the number of observations with the aim of establishing a full monitoring system by 2006.

CarboEurope will also monitor the reaction of sinks to climate change. Many researchers fear that the entire European carbon sink could be saturated by 2050 and eventually become a net emitter of carbon.

Congress is due to decide on funding for the US project, known as the North American Carbon Program, before the end of the year. If funding is secured, work will start next year to establish a monitoring network by 2008.

With policy decisions resting on the results of the two studies, some researchers think that European and US carbon research should be linked more closely. "It would be politically very unwise if carbon research programmes developed continental biases too much," says Han Dolman, a meteorologist at the Wageningen University and Research Centre in the Netherlands and chairman of the CarboEurope steering committee. ■

► <http://www.bgc-jena.mpg.de/public/carboeur>

US reiterates right to keep smallpox virus for research

Washington Stocks of live smallpox virus held at the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, should be preserved to allow research into better treatments and vaccines, the US government said last week. The Department of Health and Human Services said it wants to keep stores of the virus "until adequate medical tools are available to counter any future outbreak".

The CDC is one of only two centres known to have a stock of the smallpox virus, the other being the State Research Centre of Virology and Biotechnology in Koltsovo in central Russia. The World Health Assembly — the governing body of the World Health Organization — decided in 1996 to destroy both stocks, but agreed to a temporary retention in 1999 when the United States said that further research on the live virus was necessary (see *Nature* **398**, 741; 1999). The assembly will discuss the fate of both stocks at next year's annual meeting.

Johns Hopkins admits drug misdemeanour in Indian trial

New Delhi Johns Hopkins University in Baltimore, Maryland, has admitted that one of its scientists tested experimental cancer drugs on patients in India without proper federal or university approval, and without adequate preliminary tests in animals. The researcher has been barred from serving as principal investigator on future projects involving human subjects.

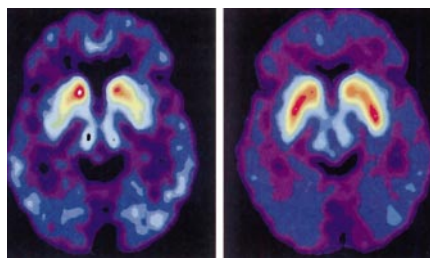
The action follows an internal investigation into an oral cancer drug trial, involving Johns Hopkins researchers, at the Regional Cancer Centre in the Indian state of Kerala between November 1999 and April 2000 (see *Nature* **412**, 466; 2001).

Johns Hopkins has now confirmed that one of the drugs had not been cleared by either the university or the US Food and Drug Administration. But it found no evidence that any patient had been harmed or that conventional treatments were delayed. The Indian health ministry has already ordered the centre to suspend all clinical trials and to reconstitute its ethical committee.

Riddle of Parkinson's side effects solved

San Diego Prospects for using of grafts of fetal tissue to treat patients with Parkinson's disease look brighter now that researchers can explain the distressing side effects reported this year in a clinical trial.

In March, Curt Freed of the University of Colorado in Denver and his colleagues



Brain wave: transplanting fetal tissue (right) boosts dopamine release in Parkinson's patients.

published a clinical trial in which 20 Parkinson's patients received transplants of human fetal tissue to replace cells in an area of the brain called the putamen (*The New England Journal of Medicine* **344**, 710–719; 2001). Cells in this area release the neurotransmitter dopamine, which is deficient in Parkinson's patients. Most treated patients aged under 60 showed significant improvement, but some were left with intractable movement disorders (see *Nature* **410**, 401; 2001).

Follow-up studies reported at last week's annual Society for Neuroscience meeting in San Diego indicate that the side effects are due to large numbers of cells being implanted into areas where dopamine function was still intact. Using brain imaging to target the transplants to areas with the greatest loss of dopamine should reduce side effects, say the scientists.

Parliament amends plans for Framework funding

Munich The European Parliament has put forward a raft of changes to the proposed sixth European Framework Programme for Research, which will run from 2002 to 2006.

The parliament last week suggested 334 changes to the 17.5-billion-euro (US\$15 billion) programme. It said that infrastructure funding should be halved to 475 million euros, while giving fusion research a provisional increase of 100 million euros. The parliament also recommended that more emphasis be placed on diseases such as malaria and AIDS, rather than newer fields such as genomics. Funds can also be used for research using human embryonic stem cells, the parliament said, as long as the research is in line with the legislation of the member states involved.

The European Commission will draft a revised version of the programme before the council of European research ministers meets again in December.

Courts blame 'slow' government for CJD

Tokyo Two district courts have found that the Japanese government and two biomedical companies could have acted earlier to prevent brain-surgery patients in Japan contracting Creutzfeldt–Jakob disease (CJD).

Seventy-six patients are known to have contracted CJD in Japan between 1979 and 1997 after receiving grafts of brain tissue imported from Germany. In two lawsuits brought by relatives of 28 of the patients, the courts found that the medical services companies that supplied the tissues — B. Braun of Melsungen, Germany, and Nihon BSS of Tokyo — should have known of the risk. The courts also found that Japan's health ministry should have acted earlier — it did not recall the tissue until 1997, a decade after the US Centers for Disease Control and Prevention banned its use.

The courts said that the companies should reach a settlement with the families, but did not specify how much compensation should be paid. The families are seeking a combined total of 2.9 billion yen (US\$23.5 million).

Weizmann picks plant man as next president



Ilan Chet: keen on links with centres abroad.

Jerusalem Ilan Chet, an agricultural scientist specializing in the biological control of plant diseases, has been elected president of the Weizmann Institute of Science in Rehovot. Chet recently completed a ten-year stint as the vice-president for research and development at the Hebrew University of Jerusalem.

Chet was founding director of the Hebrew University's Otto Warburg Minerva Center for Agricultural Biotechnology, and served for three years as the university's dean of agriculture. He has been a vigorous promoter of joint projects between researchers in Israel and overseas, fostering cooperation that generated new sources of funding for Israeli scientists. Chet replaces physicist Haim Harari, who had been in the post for 13 years.

Women close gap but it's still a man's world

Washington Female scientists have increased their representation in the US scientific community, but still lag behind their male colleagues in attaining full-time, tenure-track positions, according to a report by the US National Academy of Sciences. Women had almost no representation in the community 25 years ago, but in 1995 they accounted for around a third of new science PhDs and academic faculty in many fields, though there is wide variation.

In engineering just 7% of degrees and 5% of PhD-level jobs in 1995 went to women, whereas in the biological sciences female scientists were awarded 50% of all undergraduate degrees and 40% of all PhDs.

♦ www.nas.edu

The regeneration gap

Newts grow new legs, *Hydra* new heads. These remarkable creatures may hold clues for researchers developing human cellular therapies. But the connections are only now starting to be made. Helen Pearson reports.

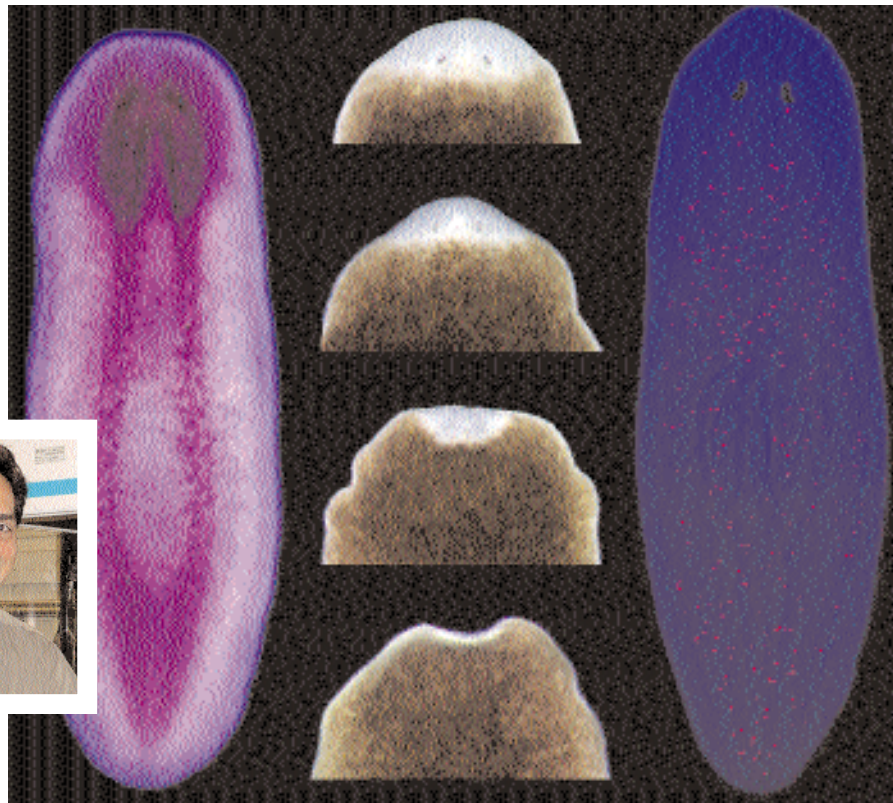
Take one flatworm, chop into 279 pieces and leave for two weeks. Feed occasionally. The result: 279 perfect new worms. The ability of flatworms, or planarians, to regrow an entire body from a handful of cells seems almost miraculous.

Salamanders, starfish, tentacle-waving polyps and zebrafish — many and varied are the organisms that can regenerate new heads, limbs, internal organs or other body parts if the originals are lost or damaged. Unfortunately, people cannot. But over the past few years, researchers studying regenerating creatures have begun to identify the genes, proteins and signalling pathways that underlie these organisms' abilities. This work indicates that the gulf between us and them is not so great. "We have the genes planarians use to regenerate their brain, muscle, their entire head," says Alejandro Sánchez Alvarado of the University of Utah in Salt Lake City.

Millions of dollars are now pouring into work on human stem cells in the hope that they can be used to regrow tissues lost to injury or disease. But Sánchez Alvarado and others who work on naturally regenerating organisms argue that molecular clues gleaned from their research might aid scientists trying to develop human cellular therapies. "We're all looking at the same kind of phenomena," says David Stocum, who studies regeneration in tadpoles of the frog *Xenopus laevis* at Indiana University–Purdue University in Indianapolis.

Regenerating organisms take one of two approaches to replacing a lost body part. Some, such as flatworms and the polyp *Hydra*, retain populations of stem cells throughout their lives, which are mobilized when needed. Unlike the 'adult' stem cells found in many of our tissues, which are thought to have a relatively restricted capacity for developing into different cell types, these cells retain the ability to regrow many of the body's tissues.

Other organisms, including newts, segmented worms and zebrafish, convert 'differentiated' adult cells — which have stopped dividing and form part of the skin, muscle or another tissue — back into stem



Wonder worms: Alejandro Sánchez Alvarado believes organisms that can regenerate body parts, such as this flatworm shown regrowing its head (centre), could aid human stem-cell research.

cells. This is known as dedifferentiation.

In both types of organism, researchers are now investigating where the cells involved get their instructions from and which genes and proteins are responsible for regeneration. They have been aided by the application of molecular and genetic techniques that are helping to turn these biological curiosities into convenient laboratory organisms (see 'Tools of the trade', opposite).

Working flat out

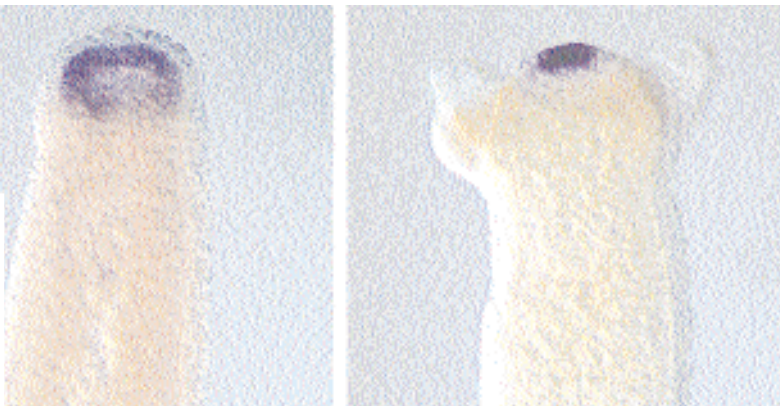
The prolific properties of planarian worms make them an ideal starting point for investigating regeneration. A centimetre-long transparent flatworm contains dormant stem cells distributed throughout its body. When damage occurs, cells near the injury site work out where they lie in the body, and carry out very specific local repairs.

Kiyokazu Agata of the RIKEN Center for Developmental Biology in Kobe, Japan, has

shown that planarian stem cells rely on signals from neighbouring damaged tissues to work out their location — and hence what repairs are needed¹. Agata and his colleagues killed the stem cells in one *Dugesia japonica* worm using X-rays. They then took a section of the irradiated worm, and grafted it upside down into a second worm, replacing the corresponding section of that worm's body.

The intact stem cells neighbouring the grafted tissue produced replacement cell types in the grafted tissue that were consistent with the reversed orientation of the inserted segment. The results convinced Agata, who is now at Okayama University, that cues from differentiated cells in the local environment are essential to instruct stem cells to regenerate correctly. He believes this tenet will also hold for mammalian stem cells: directing them to make desired tissue types will depend on telling them where they are in the body.

Thomas Holstein and his colleagues at



Body plan: Thomas Holstein found Wnt protein (blue stain) is active in *Hydra* regrowth.

the Technical University of Darmstadt in Germany, meanwhile, are among several research groups showing that the molecules responsible for telling cells their location in the polyp *Hydra* are those that perform similar jobs in higher animals. But in mammals, they are mainly active during embryological development.

Abundant in freshwater ponds and lakes, *Hydra* are hollow tubes of cells wrapped around a stomach, with a 'head' of swirling tentacles that captures food. Just like the multiheaded monster of Greek myth, which grew two new heads for every one lopped off, a *Hydra* can regenerate a new head after decapitation.

Brought to a head

Its body walls are made up of constantly dividing stem cells in which newborn cells move slowly upwards to form the stinging tentacles, downwards to form the foot, or bud off at the sides to make replica animals. "The basic processes of growth and differentiation are constantly going on," says Holstein. But to develop appropriately, the roaming cells must continuously be told where they are in the organism.

Last year, Holstein's group identified a *Hydra* member of the Wnt family of secreted proteins, which are involved in setting up the body plan in vertebrates. The researchers showed that this protein is produced just as a baby *Hydra* begins to bud away from its parent, and at the tip of a decapitated *Hydra* as its head regrows².

Holstein's team has also established that just a handful of cells from this tip is all that is needed to build a new head. The researchers decapitated several *Hydra* so that regeneration was under way, mashed the regenerating

tips into a pulp of single cells, and then allowed them to form clusters of different sizes. The clusters were added back to clumps of other body cells to find the minimum number of head cells needed to reform a head, which turned out to be about ten³.

Other researchers are similarly finding that signalling molecules or regulatory proteins involved in development in higher animals, including mammals, are also involved in regeneration in *Hydra*^{4,5}. Indeed, biologists working on the polyps believe that the minimal set of such genes commonly used to map out a growing body or limb in complex animals are present in *Hydra*. "It doesn't matter if you make a tentacle or a tooth," says Brigitte Galliot of the University of Geneva. "The question is similar: how can you make a three-dimensional structure from a pool of stem cells?"

Although much work remains to be done, this question presents an intriguing possibility. Research on *Hydra* or planarian worms might reveal how to selectively reactivate the genes and proteins that direct early mammalian development, and so induce regeneration in injured or diseased human tissues. And for researchers working on organisms that dedifferentiate adult cells before regenerating, rather than retaining populations of dormant stem cells, such connections with work on mammals are already being made.

Back to the future

Jeremy Brookes of University College London describes urodele amphibians — newts and salamanders — as "champions of regeneration". Uniquely among vertebrates, they can regrow a wide variety of lost or damaged body parts including limbs, tails, parts of the eye, and even large chunks of the heart. In each case, they achieve this by first dedifferentiating neighbouring cells.

Over the past few years, biologists interested in regenerative medicine have started trying to induce similar developmental reprogramming in mammalian cells⁶. If they could transform any adult cell into a stem cell capable of division and repair, it could bring

Tools of the trade

The regenerative powers of *Hydra* polyps have fascinated experimental biologists for more than 250 years. Yet until recently, the processes involved remained unexplained. "The classical literature is chock-full of experiments that need a molecular explanation," says Alejandro Sánchez Alvarado of the University of Utah in Salt Lake City.

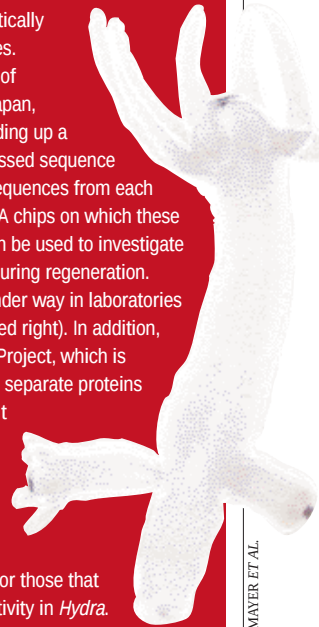
One problem is that regenerating organisms do not lend themselves to experimental genetics. *Hydra*, for example, have a genome as big as that of humans, and refuse to have sex in the lab, making crosses to study the inheritance of traits impossible.

But over the past few years, these extraordinary organisms have been undergoing a renaissance, thanks to the discovery in other species of genes considered to be good candidates for having a role in regeneration, and the advent of techniques for high-throughput analyses of gene activity.

Now, efforts are under way to give regenerating organisms their own genetic toolkits. Two years ago, while he was at the Carnegie Institution of Washington in Baltimore, Sánchez Alvarado's team showed that RNA interference, in which double-stranded molecules of RNA are used to block a specific gene's activity, could be refined for use in flatworms¹⁴. This will allow researchers to track down the genes involved in regeneration by systematically disabling likely candidates.

With Kiyokazu Agata of Okayama University in Japan, Sánchez Alvarado is building up a library of flatworm expressed sequence tags (ESTs), short DNA sequences from each of the worms' genes. DNA chips on which these ESTs are immobilized can be used to investigate which genes are active during regeneration.

Similar efforts are under way in laboratories working on *Hydra* (pictured right). In addition, there is a *Hydra* Peptide Project, which is using chromatography to separate proteins produced by the polyps. It then screens for protein fragments that might be involved in the signalling pathways that direct development and regeneration by looking for those that alter patterns of gene activity in *Hydra*. The project has so far identified nearly 1,000 candidate peptides¹⁵.



You could create stem cells when you need them, where you want them.

about a medical revolution. So far, they have not had much success, but recent crossovers from work in newts into research on mice have generated a frisson of excitement.

Inspired by newts' capabilities, two research groups have attempted to induce dedifferentiation in mouse muscle. As differentiated adult muscle forms, individual cells stop dividing and fuse together into a myotube, a long muscle fibre containing many cell nuclei. Newt myotubes can reverse

▶ this process, giving rise to stem cells that can regenerate a limb.

A team led by Peter Schultz of the Scripps Research Institute in San Diego, California, synthesized large libraries of purines, a class of small molecules that commonly bind to proteins and influence a variety of cellular processes. They screened for purines that reversed the differentiation of mouse myotubes, and pulled out one, which they dubbed myoseverin, that caused cells to split off and begin to divide⁷. The compound binds to a component of the myotube's internal skeleton, and might break down the fibre's internal structure.

Muscle switch

Mark Keating of Harvard Medical School in Boston and his colleagues have concentrated on a gene called *msx1*, which encodes a protein that controls the activity of other genes in muscle, and is active both in the regenerating stump of a severed newt leg and in embryonic mouse limbs.

The researchers genetically engineered mouse muscle cells so that the *msx1* gene could be switched off by administering an antibiotic. The team then allowed the cells to develop into myotubes. When the antibiotic was removed, activating *msx1*, genes that are normally expressed during the differentiation of mouse myotubes were switched off in

reverse order, and individual cells began to break off from the myotubes and divide⁸. Furthermore, these dedifferentiated cells, when plied with appropriate growth factors, could form fat, bone and cartilage as well as muscle. "They have the ability to go back in time," says Keating.

Keating and Brookes have since been working out what normally stops mammalian muscle from undergoing dedifferentiation. Brookes believes that the block arises from a failure to activate the appropriate cellular signalling pathways inside mammalian myotubes. But by merging newt and mouse muscle cells to form hybrid myotubes, his team has found that the mouse nuclei can be induced to start copying their DNA — a prelude to cell division⁹.

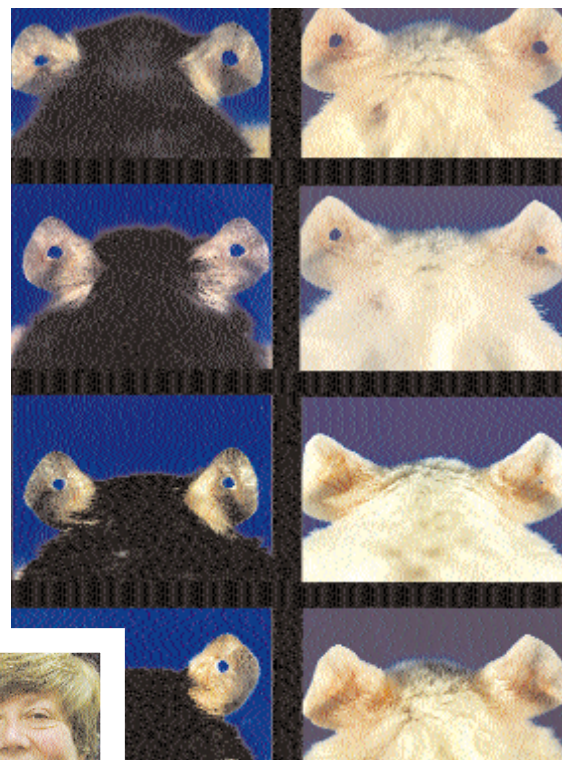
Keating, meanwhile, has found that liquidized extracts from regenerating newt myotubes can trigger the dedifferentiation of mammalian myotubes¹⁰, suggesting that a protein present in a newt limb stumps triggers the dormant pathways. "It may be all that's required to get regeneration," Keating speculates. If that elusive protein can be tracked down, he suggests, "you could create stem cells when you need them, where you want them."

Some evidence has emerged that certain strains of mouse are unusually predisposed to regeneration. Ellen Heber-Katz of the Wistar Institute in Philadelphia punched 2-millimetre holes in the ears of members of a laboratory strain called MRL — a standard technique used to tell individuals apart. Three weeks later, there were no holes and no scars¹¹. Although such healing is usual in rabbits, it had never before been seen in mice. "I was virtually jumping up and down," says Heber-Katz. "Clearly these mice can do things that others can't."

Growth potential

Since stumbling on this quirk, Heber-Katz has studied the MRL strain's regenerative capacities. When one chamber of their heart is frozen, the mice take only two months to recover normal working heart muscle¹². In other strains, the damaged muscle does not recover. MRL mice appear to form less scar tissue than other strains, and this may be their secret. Scar tissue inhibits regeneration, Heber-Katz believes. She suspects that, in MRL mice, enzymes that break down scar tissue may be over-active. Although her team has shown in studies of the trait's inheritance that several genes must underlie MRL's regenerative abilities¹³, these have yet to be identified.

Despite the upsurge of interest in mammalian stem-cell research and regenerative medicine, relatively few researchers working in this area seem aware of the poten-



Hole picture: the amazing healing abilities of MRL mice (right) have excited Ellen Heber-Katz.



tial applications of findings that are emerging from regenerating organisms. "I don't think people have made the leap," says Stephen

Johnson, who studies the regeneration of zebrafish fins at the Washington University School of Medicine in St Louis, Missouri.

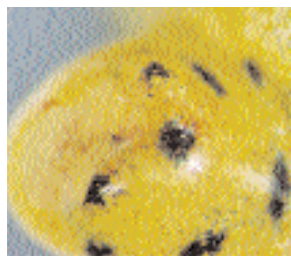
Diane Krause of Yale University, who works on the bone marrow stem cells that give rise to blood cells, agrees that researchers trying to develop the field of regenerative medicine should start paying more attention to the work of biologists studying newts, *Hydra* and other champions of regeneration. "Do I follow them? No. Should I? Yes."

Helen Pearson works in Nature's science writing team.

1. Kato, K., Orii, H., Watanabe, K. & Agata, K. *Dev. Biol.* **233**, 109–121 (2001).
2. Hobmayer, B. et al. *Nature* **407**, 186–189 (2000).
3. Technau, U. et al. *Proc. Natl Acad. Sci. USA* **97**, 12127–12131 (2000).
4. Smith, K. M., Gee, L. & Bode, H. R. *Development* **127**, 4743–4752 (2000).
5. Guachat, D. et al. *Proc. Natl Acad. Sci. USA* **97**, 4493–4498 (2000).
6. Aldhous, P. *Nature* **410**, 622–625 (2001).
7. Rosania, G. R. et al. *Nature Biotechnol.* **18**, 304–308 (2000).
8. Odelberg, S. J., Kollhoff, A. & Keating, M. T. *Cell* **103**, 1099–1109 (2000).
9. Velloso, C. P., Simon, A. & Brookes, J. P. *Curr. Biol.* **11**, 855–858 (2001).
10. McGann, C. J., Odelberg, S. J. & Keating, M. T. *Proc. Natl Acad. Sci. USA* **98**, 13699–13704 (2001).
11. Clark, L. D., Clark, R. K. & Heber-Katz, E. *Clin. Immunol. Immunopathol.* **88**, 35–45 (1998).
12. Leferovich, J. M. et al. *Proc. Natl Acad. Sci. USA* **98**, 9830–9835 (2001).
13. McBrearty, B. A., Clark, L. D., Zhang, X.-M., Blankenhorn, E. P. & Heber-Katz, E. *Proc. Natl Acad. Sci. USA* **95**, 11792–11797 (1998).
14. Sánchez Alvarado, A. & Newmark, P. A. *Proc. Natl Acad. Sci. USA* **96**, 5049–5054 (1999).
15. Bosch, T. C. G. & Fujisawa, T. *BioEssays* **23**, 420–427 (2001).



Back on its feet: Mark Keating has identified a gene involved in limb regrowth in newts.



Pushing the physics envelope

Canada's new Perimeter Institute is planning to apply the risk-taking approach of venture capitalism to the pursuit of theoretical physics, says David Spurgeon.

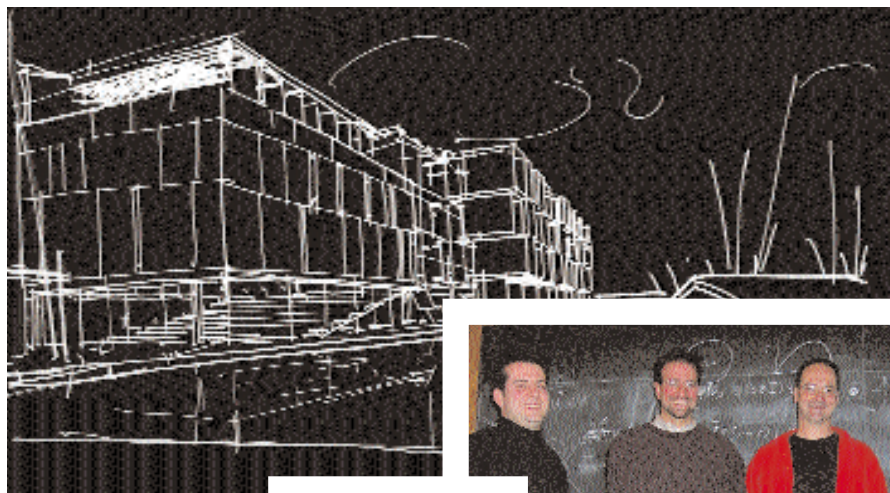
When Mike Lazaridis was a boy growing up in Windsor, Ontario, he was fascinated by physics. But when he went to the University of Waterloo he chose to study engineering, for the simple reason that there weren't enough job opportunities in physics.

Now, having made his fortune as president of a company called Research In Motion, which makes wireless Internet devices, Lazaridis has returned to his first love. In October 2000, he announced a gift of Can\$100 million (US\$63 million) to establish the Perimeter Institute for Theoretical Physics, supplemented by Can\$20 million from two other executives at Research In Motion (see *Nature* **408**, 125; 2000).

One year later, as the institute's first recruits start settling into their temporary quarters in Waterloo's old post office, they are hoping to apply Lazaridis's entrepreneurial spirit to some of the most difficult questions in physics. "The institute was explicitly created to encourage young, innovative scientists in their research prime to pursue issues that are often deemed 'too hard' and 'only to be done after procuring tenure,'" says Howard Burton, its executive director. Perimeter scientists will tackle notoriously thorny problems in areas such as quantum gravity, quantum information theory, elementary particles and cosmology.

For Lee Smolin, the risk-tolerant environment promised by Lazaridis and Burton was instrumental in his decision to join the institute. In fields such as his speciality of quantum gravity, Smolin believes scientists can learn from high-tech industry, where venture capitalists typically look for a 10% success rate. "It would be much better if we could work in an environment in which we're perfectly honest about the risks we take," says Smolin, formerly of Pennsylvania State University and Imperial College London. "One of the things that particularly excites me about Perimeter is the willingness to not only accept such risk, but actually to embrace it."

Smolin argues that most academic physics departments also suffer from narrow specialization — usually including representatives of only one theoretical approach to a particular problem. In trying to devise a theory of quantum gravity, for instance, there are two main approaches. The first, string theory, aims to unite gravity with the other forces of nature by viewing all fundamental particles as infinites-



imal vibrating strings. The second seeks to make the geometrical view of space-time proposed in Einstein's theory of general relativity compatible with quantum mechanics by invoking similarly tiny geometrical entities called loops.

"Unfortunately, there are places where one approach or another is dominant," says Smolin. But at Perimeter, he says, both will be represented equally. And the institute is looking for recruits from either camp who are sympathetic to the views of the other — Rob Myers, for instance, a Perimeter Institute string theorist formerly of McGill University in Montreal, also has a strong background in general relativity.

Thinking space

Construction of a permanent home for Perimeter should begin next February on a site donated by the city of Waterloo, overlooking a lake and park a few minutes' walk from the city's university. The architects plan to achieve a harmonious balance between private spaces for quiet contemplation and calculation, and public areas for discussion.

Interaction is a watchword for the Perimeter Institute. "The areas we're working on at Perimeter all have a common theme, in that they're looking at the foundations of physics," says Myers. "The potential for interactions between all these diverse fields is the important aspect." To help this, the institute is trying to hire theorists with a polymath bent, such as Fotini Markopoulou, who comes to Perimeter from the Albert Einstein Institute in Potsdam, Germany, and has worked in quantum gravity, quantum cosmology,



Physicists (from left) Michele Mosca, Howard Burton, Rob Myers and Raymond Laflamme have all joined the Perimeter Institute, founded by Mike Lazaridis (inset).

causality, quantum computing and logic.

Currently, the institute has seven members — three permanent staff and four postdocs. Within seven years, the new building should house some 40 resident physicists and postdocs as well as about 30 visitors and associate members. Its first associate members are Raymond Laflamme and Michele Mosca, experts in quantum information theory at the University of Waterloo.

"They're off to a good start," observes University of Oxford mathematician Roger Penrose, a member of the institute's scientific advisory board. "Of course, it depends on how the recruitment goes from now on, but I expect Perimeter to do well."

The institute has already launched a high-school outreach programme, and recently initiated a public lecture series. By creating jobs for theoretical physicists and reaching out to schoolchildren, Lazaridis hopes that the institute will encourage budding physicists to stick with the subject, rather than being diverted as he was along other — admittedly more lucrative — paths. "Perimeter is an attempt by me and my colleagues to say physics really matters and that there are people in the community who really understand its importance and are willing to step up and fund it," he says.

David Spurgeon writes for *Nature* from Montreal.

► www.perimeterinstitute.com

R. J. EPP

IMAGE POWER

Scientific data must be made available to all

An internationally binding regulation should be the first step towards securing vital data.

Sir—Recent correspondents (see for example, K. Alverson and M. Eakin, and M. Boulter, *Nature* **412**, 269; 2001) emphasize the importance of general repositories for scientific data. They point out the lack of appropriate financing for scientific data management and the need for thematically related scientific-information systems to be networked, which is technically feasible.

The current problem resides deeper. To enact global agreements such as the Kyoto protocol, for example, decision-makers are required to pursue policies that cover many different, yet plausible, estimates of the likelihood of alternative, future climate development. These climate-change scenarios are frequently derived from climate models, whose correctness can be measured only by access to raw data.

The necessary database infrastructure was created as early as the 1950s, when the International Council of Scientific Unions set up the World Data Center (WDC) system (www.ngdc.noaa.gov/wdc). These centres are for the international exchange of solar, geophysical and related environmental data on a long-term basis, and are to assist principal investigators (PIs) in broad data management. Yet there are no international regulations requiring scientists to store results as raw data and accompanying meta-information in this or any other publicly accessible archive.

Some scientific journals and funding agencies encourage PIs or authors to submit raw data or support data sharing. Yet the US National Science Foundation (NSF) is the only funding organization that insists on this. In its information on submitting climate-related proposals for its Earth System History Program (www.nsf.gov/pubs/2000/nsf0011/nsf0011.html), the NSF says: "Each proposal must adhere to the USGCRP [US Global Change Research Program] data management policy ... and the policies applying to recipients of Federal funding in the geosciences. Unless otherwise specified in the proposal, the [PI] will be responsible for ensuring that all data generated by the funded project will be documented and submitted to the World Data Center".

Our example of climate-change models is but one of many where raw data are needed for accurate scientific assessment in the long term. An internationally binding regulation that adheres to WDC principles is required. It should guarantee that all scientific data are archived and freely available. Because some PIs still refuse to archive data individually in

appropriate databases such as the WDC, still refuse to make their published data publicly available as raw data, and still ignore the potential benefit from networking through WDCs, additional financial and temporal effort must be spent to gain access to these data.

Until a free-access global repository can be set up, this may be the best first step. The long-term goal must be to secure the world's stock of valuable data, and to

overcome practical difficulties in so doing.
Nicolas Ditter*, **Michael Diepenbroek†**,
Hannes Grobe‡

**SINOPS Project Office France, Institut Universitaire Européen de la Mer, UMR 6539 - LEMAR, Place Nicolas Copernic, Plouzané F-29280, France*

†World Data Center for Marine Environmental Sciences, Klagenfurterstr., Bremen D-28359, Germany

‡PANGAEA Network for Geological and Environmental Data, Alfred Wegener Institute for Polar and Marine Research, Bremerhaven D-27515, Germany

Speak out against wrong done in your name

Sir—Edgar Pick in his Correspondence "Science is universal, not part of any religion" (*Nature* **414**, 249; 2001) is right in pointing out that simply trying to engage with 'Islamic' science is not a way for western (presumably 'non-Islamic') scientists to contribute to the struggle against terrorism. However, he is wrong in saying "the evil of killing people ... hardly requires a restatement" by Islamic scholars.

The terrorists who flew planes into the World Trade Center may very well be "an international gang of well-financed criminals" but they apparently carried out these acts in the name of Islam. If a group of deranged scientists began slaughtering innocent people in the name of cell biology, then I would be the first to stand up and say "Hang on a minute, not all cell biologists adhere to this view".

To the millions in the non-Islamic world who have neither Pick's education nor his knowledge of Islam, the statements of condemnation by Islamic leaders and scholars provided essential reassurance and may have helped deflect retributive acts by the more thuggish elements of our society.

Stephen E. Moss

Division of Cell Biology, Institute of Ophthalmology, University College London, Bath Street, London EC1V 9EL, UK

Chimaeric mice on the road towards stem cells

Sir—In his engaging Commentary "IVF and the history of stem cells"¹ Bob Edwards gives me undue credit in asserting that I produced the world's first mouse chimaera. Unquestionably, it is Krzysztof Tarkowski who should be credited with this. He aggregated mechanically denuded

mouse morulae in pairs to obtain giant blastocysts, some of which developed normally to term following transfer to uterine foster-mothers². Although perinatal losses were inexplicably high, seven of nine available young showed a mixture of pigmented and unpigmented cells in their eyes, in accordance with the genotypes of the morulae that were combined³. Subsequently, Beatrice Mintz introduced various technical improvements that enabled chimaeric mice to be produced routinely⁴.

My contribution was to show that similar chimaeras could be obtained by transplanting cells between blastocysts⁵. This was a refinement that enhanced the value of such organisms for studying development⁶ and was later adopted as the standard way of obtaining germline transmission of genetically modified embryonic stem cells.

Richard Gardner

Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

1. Edwards, R. G. *Nature* **413**, 349–351 (2001).
2. Tarkowski, A. K. *Nature* **190**, 857–860 (1961).
3. Tarkowski, A. K. *NCI Monograph* **11**, 51–71 (1962).
4. Mintz, B. *Science* **138**, 594–595 (1962).
5. Gardner, R. L. *Nature* **220**, 596–597 (1968).
6. Gardner, R. L. *BioEssays* **20**, 168–190 (1997).

The perils of peer review

Sir—Your News feature "Peers under pressure" (*Nature* **413**, 102–4; 2001) on the hoary old chestnut of peer review reinforces my decades-old comparison of this ritual to the Latin mass. Obviously many (Protestant?) leaders, including most of the best-known scientists such as Nobel laureates, regard peer-review as a great hindrance to good science (the gospel?). Many excellent journals (churches?), such as the *Proceedings of the Royal Society* and *Proceedings of the National Academy of Sciences* managed in my opinion very well without it for a long time. An enormous amount of the best science has been and is

being run without benefit of this rubric, as is the worldwide patent system.

The allusion in your Feature to Paul Chu and the 1:2:3 superconductor paper — the contents of which had clearly reached rival research groups before publication, and in which “ytterbium” was changed to “yttrium” at the last minute — points to but one of the very minor ineradicable defects of the peer-review system, that we can confirm. At the time that paper was submitted, several of its authors, including one of us (J. R. A.) who had made the very first samples of 1:2:3 at the University of Alabama, had a conversation with his Alabama colleagues and Dr Chu of the University of Houston, at which the wise decision was made to substitute Yb for Y in the text. The galley proof (still in J.R.A.’s possession) shows Yb, not Y, in all four places where it appears. The full chemical name of ytterbium — and, subsequently, yttrium — was conveniently omitted from the text of the paper so that the amendments before final printing could be confined only to the symbols. This stratagem was also alluded to in the Materials Research Society videotapes on the history of the 1:2:3 discovery, organized by one of us (R.R.). Everyone except the true believers knows that it is your nearest competitors (adversaries?) who often ‘peer’ review your paper. Hence, you must protect yourself by this and other subterfuges, like proposing work you have just completed.

Yet this is but a minor defect in the peer-review system. The enormous waste of scientists’ time, and the absolute, ineluctable bias against innovation, are its worst offences.

‘Review by competitors’ is an all-too-accurate description of this system, wreaking devastation on papers and proposals in science. Financial and self-interest disclosures, such as competing for the same funds, should surely now be required of peers.

More: where is the evidence for any benefits from peer review? Recently, *Nature* (*Nature* **412**, 751; 2001) and most medical journals have been forced to require financial disclosures by authors to deal with the fact that peer review could do nothing to avoid the widely acknowledged contamination of the literature. This was illustrated by papers published about the drugs Vioxx and Celebrex — just one example picked up by the secular press (*Washington Post*, 5 August 2001; *Wall Street Journal*, 22 August 2001) in front-page stories after a paper in the *Journal of the American Medical Association* (*J. Am. Med. Assoc.* **286**, 8; 2001) reported that these popular painkillers carried a risk of cardiovascular problems. The newspapers reported that cardiovascular risks from

Celebrex had been simply “omitted” from an earlier paper (*J. Am. Med. Assoc.* **284**, 10; 2000). All sixteen authors of the earlier paper — including faculty from eight universities — were either employees, funded by or consultants of the manufacturers. What use was peer review here?

Finally, *Nature* should not repeat the old canards such as: “despite the problems thrown up by peer review, no serious alternative has yet been proposed”. Nonsense. They have not only been proposed but have been in regular use worldwide for a very long time. The users include the world’s largest research agency, the US Department of Defense, and industrial research worldwide.

Rustum Roy*, **James R. Ashburn†**

*The Pennsylvania State University,
102 Materials Research Laboratory,
University Park, Pennsylvania 16802, USA
†Elmco, 60 Technology Drive, Building 1, Suite N,
Huntsville, Alabama 35805, USA

Carrots, not sticks, give best ‘quality assurance’

Sir — I was interested in your News story “Proposed scheme will scrutinize student supervisors” (*Nature* **413**, 761; 2001) describing the plans by the Biotechnology and Biological Sciences Research Council (BBSRC) and the Higher Education Funding Council for England (HEFCE) to improve training and standards for the supervisors of postgraduate students in the United Kingdom.

Raising the standards for postgraduate student supervision is a good thing. However, the rush by government and other higher-education bodies to introduce ‘quality assurance’ wherever conceivable, with initiatives such as the Research Assessment Exercise and the Institute for Learning and Teaching, appears little more effective in alleviating the real problems than sticking a plaster over a gangrenous wound.

The acknowledged difficulty in enrolling well-qualified postgraduate students in the United Kingdom — and the reason so many PhDs do not seek employment in their subjects — probably stems from the lack of an equitable career structure in higher education. I cannot think of another career where the most experienced practical practitioners, postdocs, live from short-term contract to short-term contract and risk being priced out of the job market when they have acquired years of experience.

This situation needs to be urgently remedied. Introduce a nationally recognized training scheme which allows research trainees to build a recognized

skills portfolio. This would help them secure jobs as they progress through their career — even if this means reducing the number of PhD places. Training could include skills such as grant writing, staff management, project management and writing papers, in addition to traditional technical hands-on abilities.

Even when postdoctoral trainees do manage to get academic positions, they find that, since the late seventies, academic salaries have fallen far below those of comparable professions and of the private sector. Students graduating with an undergraduate degree and saddled with debts of, on average, £16,000 (US\$22,600) are hardly likely to consider further years of penury and job insecurity worth the risk. Unless the government has the foresight to address these overarching problems, the future of the higher education sector and the intellectual skills base of the United Kingdom looks bleak. In the end you get the system that you pay for.

It would also be of help if carrots were used instead of sticks. How about HEFCE giving a pay rise to staff who do meet their new training standards? I would also be intrigued to know if this new spirit of ‘above-board correctness’ extends to paying realistic fees for external examiners of PhD theses. Currently the average fee is £120 to examine a PhD thesis; the Association of University Teachers recommends a fee of £450, which is not unreasonable. Try asking a management consultant to read about 250 pages, consult libraries and travel to cross-examine someone for this fee. The last time that I heard of an external examiner asking for £450, the candidate’s supervisor was told to find another examiner who would accept the lower fee.

Roddie McKenzie

Department of Dermatology, University of
Edinburgh, Lauriston Building R1E, Edinburgh
EH3 9YW, Scotland, UK

Motion and meaning

Sir — In his News and Views Feature “Where drunkards hang out” (*Nature* **413**, 686–687; 2001), Ian Stewart states “Brownian motion is no longer important in its original physical context”. I beg to differ. Analysis of diffusion as brownian motion is central to understanding the physics of diffusion-controlled biochemical reactions, such as ion permeation through protein channels, to take one example.

Eric Jakobsson

Department of Molecular and Integrative
Physiology, National Center for Supercomputing
Applications, University of Illinois, Urbana,
Illinois 61801, USA

Making it to the top

Key steps to becoming a physicist of the very first rank.

Great Physicists: The Life and Times of Leading Physicists from Galileo to Hawking

by William H. Cropper

Oxford University Press: 2001. 514 pp. \$35

David Bodanis

It's easy to be an exceptionally good physics student. There are several every year who graduate with top credentials from Caltech or Imperial or Cambridge. But track them down, as the years go on — how many will enter the history books as achieving anything of the very first rank? William Cropper has written a history of modern physicists that along the way shows how the very best made it. These are some of the key steps...

Find a great mentor

This is crucial for all but those in the Einstein league. With a great mentor you learn the latest tools, you're exposed to the latest research and you pick up a huge number of 'implicit' tools that are never put into the standard literature, but which are indispensable for coming up with top work of your own. Seeing your mentors nestled snugly in a tight hierarchy is also useful. It's a long haul to the top, and youngsters in their twenties know they're forgoing a lot of income and status by serving these apprenticeships. It's easier to put up with all that if they know great rewards await.

Overthrow that mentor

Oh, you don't have to yell and scream at the mentor. You might even pretend to be polite, and you can stay deferential at conferences and all the rest. But you do have to destroy them, otherwise the best you can hope to be remembered for is good work in the tradition of X. Werner Heisenberg loved Arnold Sommerfeld, but Heisenberg's work on quantum mechanics undercut everything Sommerfeld had taught.

Some mentors make that easy. They're so scared of the bright students they've brought on that they turn against them. This is what Sir Humphry Davy did when Michael Faraday had his first great result on electromagnetic induction. It left Faraday feeling shocked, bereft — but also freed him to do more top work on his own.

Find the right topic

It's easiest if this comes from an inner goal. A surprisingly large number of physicists have had as their lurking goal a belief in the underlying unity of nature. They can't really justify it, and are often even embarrassed to talk about it, but it's superb at guiding and motivating them.



Dearest enemy: Humphry Davy turned on his star pupil Faraday when he began to shine too brightly.

Another successful way to find a good topic is to skim the past, to find areas that everyone around you thinks have long since been completed. If they're wrong, you have a fresh, ripe topic. Einstein was profoundly motivated by reaching back to read Immanuel Kant and David Hume — their questions about causality and how we justify our knowledge were still alive, and had only been skimmed over by Einstein's contemporaries.

Faraday did a similar thing. In the early nineteenth century everyone 'knew' that empty space was not filled with invisible goblins or any other entities. Hadn't Isaac Newton proved that? But Faraday looked up some of Newton's minor letters, and saw that Newton himself had had some doubts about this.

Maybe empty space wasn't that empty after all — Faraday's concept of the field was, after a good deal of work, a direct result.

Develop your own tools

This is crucial, for you need a distinctive edge. (Again, remember that most top-rank physicists are *not* overwhelmingly smarter than others who remain at the second rank or below.) If 100 very smart people are trying to understand elementary particles and all are

using similar differential equations, who can possibly stand out? But now imagine that one of those 100 starts using group theory. He's suddenly 'smarter', or at least able to see more deeply. The other 99 are left behind.

The one weakness with this method is that you have to be sure it's the right new tool. Simply picking an arbitrary one from some unsuspected field is not going to work. One way to avoid this problem is to transpose tools from one field to another right at the moment you need them. Richard Feynman was good at this, and Enrico Fermi was exceptionally good. When he was studying interstellar clouds, he'd recognize that the magnetic lines of force threading through them somehow resembled the vibrations of a crystal lattice. The tools that had been built up for lattices could then be applied to galaxies. (When attempting such transpositions, you'll always need to check the result, but at least you have a fresh start.)

Another way to avoid using the wrong tool is to be strongly aware of the level at which you are applying your tools. Several researchers early in the twentieth century felt that there were problems with the electrodynamics of moving bodies, but most tried

to solve them using standard dynamics. Einstein took a clue from the second law of thermodynamics instead. Although the second law can give lots of detailed results, it can also serve as a higher-level 'judgement machine', telling you whether an entire type of activity is allowable or not. The result, of course, was the postulates of special relativity.

Relax in your own way

Physics is hard, and relaxation is crucial. Max Planck and Werner Heisenberg played the piano at near concert-level. They loved immersing themselves in music, not just for its beauty, but because it suggested the underlying truths that were still waiting to be discovered. Religion has been a deep source for many others — such as Abdus Salam and Faraday — and there are also a few who've taken the Schrödinger approach. Erwin Schrödinger had a tolerant wife, a charming manner and as many affairs as Wolfgang Pauli had tantrums. This presented problems in Anglo-Saxon countries. As the biographer Walter Moore put it: "It was bad enough [for Schrödinger] to have one wife at Oxford — to have two was unspeakable."

Most frequently, though, it's simply conversation: honest, quick-firing conversation with the handful of friends who can lift you to a higher level. This insight is what William Cropper has provided, in a book that has such excellent biographical information,

mixed in with fully accurate science, that it can serve both as a good guide for top secondary-school students or beginning university students, and as a wise collation of advice and experience for working physicists at all levels beyond.

David Bodanis (e-mail: davidbodanis@compuserve.com) taught the 'Intellectual Toolkit' course at the University of Oxford for many years. His most recent book is *E = mc²: A Biography of the World's Most Famous Equation* (Pan Macmillan).

Costs and benefits of inflicting pain

Defending Animal Rights

by Tom Regan

University of Illinois Press: 2001. 224 pp.

\$24.95, £18.50

Why Animal Experimentation Matters: The Use of Animals in Medical Research

edited by Ellen Frankel Paul & Jeffrey Paul

Transaction Publishers: 2001. \$49.95 (hbk),

\$24.95 (pbk)

William H. Shaw

In *Defending Animal Rights*, Tom Regan maintains that animal experimentation is immoral and should be halted. The ideas and

arguments of philosophers such as Regan have helped to galvanize the animal-rights movement. But although the regulation of animal research has increased, extremists still express their dissatisfaction with it by harassing scientists, vandalizing laboratories and releasing test animals from their cages.

The contributors to *Why Animal Experimentation Matters*, by contrast, deplore animal activism and reject the idea of animal rights. Its editors do not strive for balance, but seek to show that animal experimentation is vitally important to human welfare and therefore morally permissible. Most of the essays are by scientists who argue that animal experimentation has contributed crucially to medical and scientific progress, and offers our best hope for understanding and combating many deadly or incapacitating diseases. But lists of scientific breakthroughs resulting from animal testing and experimentation will not assuage those critics, including some scientists, who believe that the overall benefits of animal experimentation are exaggerated.

Some of the authors in *Why Animal Experimentation Matters* worry that excessive concern for animal welfare can hamper research, and they write as if there have never been needless or abusive experiments on animals. These days, Institutional Animal Care and Use Committees (legally mandated in the United States) typically urge experimenters to 'replace' animal experimentation with other research methods where possible, to 'reduce' the number of animals used, and to 'refine' the research so as to minimize pain and suffering (the 3R approach). Experimenters are also encouraged to substitute lower animals for higher ones. Has such regulatory control eliminated genuine evils? Has it retarded scientific progress? This book neglects such questions.

The scientists writing in *Why Animal Experimentation Matters* believe that humans matter more than animals do, but their moral arguments are often superficial. Placing emphasis on the benefits we would lose if we gave up animal research does not prove that animals have no moral rights or that their interests are inherently less valuable than our own. Neither our superior cognitive abilities nor the fact that animals treat one another badly settles the question of our duties to them. And one reads with embarrassment two evolutionary biologists decrying animal rights as a "maladaptive philosophy" because it fails to promote the interests of our species, and debunking animal advocates as "adaptively unfit" because they tend to be childless (based on a survey of participants in a 1990 rally).

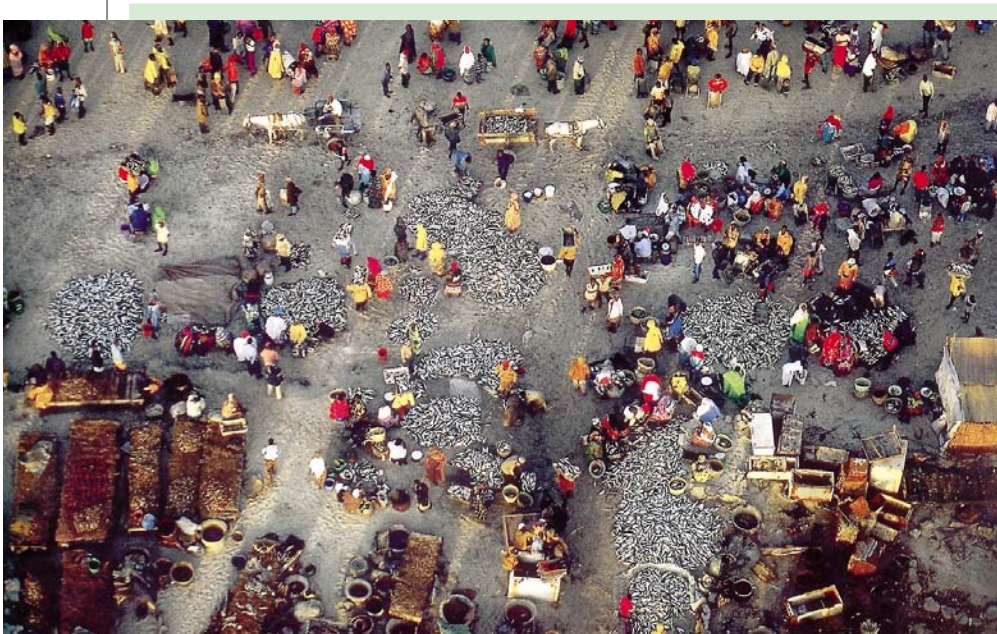
By contrast, the philosophers contributing to the book grasp the deeper, thorny ethical issues. And, unlike the scientists, they don't mind raising problems for the home team. Thus, Baruch Brody proposes a thought

Rare and unusual



The Usambara three-horned chameleon (*Chamaeleo deremensis*) from Tanzania becomes spotted with black when irritated. It is one of the

420 or so reptiles illustrated in *A Field Guide to the Reptiles of East Africa* by Stephen Sprawls et al. (Natural World, £29.95).



Around the world in 365 days

The vivid colours of a fish market in Saint-Louis, Senegal, one of Yann Arthus-Bertrand's photographs in *The Earth from the Air: 365 Days* (Thames & Hudson, £24.95). His powerful

images on many different scales range from the polar wastes to the tropics, giving a new view of natural and human landscapes — from the 'stone forest' of Madagascar to the Rio slums.

experiment that forcefully challenges the assumption that animal pain is intrinsically less important than human pain, and R. G. Frey argues that the same reasoning that justifies experimenting on animals (which he supports) also justifies experimenting on members of our own species.

Regan, on the other hand, is an abolitionist. He does not want to regulate or reform animal experimentation, but to end it altogether. He rejects the utilitarian approach of philosopher Peter Singer. Singer believes that animal testing and experimentation are wrong because the price paid by the animals outweighs the likely benefits to humans. Singer does not contend that animal lives are equal in value to our own or that they have rights. What he does say is that, to act morally, humans must take animal interests into account, giving those interests as much weight as the equivalent interests of humans.

Regan faults utilitarianism for permitting us to sacrifice the interests of some for the benefit of others and, more specifically, for failing to rule out animal experimentation in principle. He believes that some animals are the "subject of a life". By this he means that they "bring the mystery of a unified psychological presence to the world". These animals desire, remember, feel emotion and act intentionally. Taken together, such capacities mean that their lives have innate value. In kantian terminology, these animals are ends in themselves — just as humans are — and, accordingly, they have a right to life, liberty and bodily integrity. More generally, they

have a right to be treated with respect.

If all subjects of a life have these basic rights, and if, as Regan believes, many non-humans are subjects of a life, then experimenting on these animals is morally impermissible. To point to the costs of forgoing such experimentation, as *Why Animal Experimentation Matters* does, misses the point. But is Regan's reasoning sound? Regan argued his case at length in his 1983 book, *The Case for Animal Rights* (University of California Press). In several of the essays collected in *Defending Animals Rights*, he restates his position and addresses his critics. But the philosophical debate is bound to rage on. ■

William H. Shaw is in the Department of Philosophy, San José State University, 1 Washington Square, San José, California 95192-0096, USA.

Biotechnology retrospective

The Green Phoenix: A History of Genetically Modified Plants

by Paul F. Lurquin

Columbia University Press: 2001. 240 pp.

\$50, £33.50 (hbk), \$25, £17 (pbk)

F. C. Botha

Ask any student of biotechnology today when the first foreign DNA was transferred to plants, and the answer will invariably be 1984. Very few will recall that it all started

with the innovative and tenacious work in Lucien Ledoux's laboratory in Belgium in the late 1960s. And it would be interesting to know how many plant molecular biologists have read the two pioneering publications of Ledoux and Huart, which claimed to report the integration of foreign DNA into the genome of barley. Paul Lurquin's book appropriately reminds us of the major difference between the theory of the scientific method and the way it translates into practice. Certainly, it is not as simple as formulating a hypothesis and then automatically verifying it. All those many failures — and often erroneous interpretations — are seldom told to newcomers to the field.

In the late 1960s, the concept of horizontal DNA transfer — the transfer of genes between species — was inconceivable, as this violated the accepted idea of slow, mutation-driven evolution constrained by sexual barriers. Without good models or detection systems, the early pioneers in this field had to persuade a highly sceptical scientific community of the merits of their hypotheses. It is almost unimaginable that they had to rely on a very basic technique, density centrifugation, to demonstrate gene transfer in plants. Today's students have the benefit of sensitive modern techniques, and yet often fail to obtain good evidence for the stable integration of foreign DNA into plant genomes. They will certainly appreciate the enormous task that early researchers faced in convincing others of this phenomenon.

In the very early days of plant genetic engineering, the small flowering plant *Arabidopsis* was already the laboratory workhorse for the Ledoux team. In 1974, they stated that they had achieved the successful complementation (restoration to normal function) of a vitamin B1 mutation by the transference of genetic material from bacteria. In the same period, D. Hess in Germany claimed to have engineered a change in flower colour through horizontal gene transfer. Both pieces of research were heavily criticized and their accuracy was questioned. But a major turning point came with the discovery in Germany and the USA that genes from the bacterium *Agrobacterium* can transfer naturally to plants. The book accurately describes the very rapid developments that followed this discovery, culminating in the conclusive demonstration of horizontal gene transfer to plants in 1984.

Looking back at the turbulent first two decades of plant genetic engineering, it is difficult to understand how the researchers involved persisted with their ideas. It is inspiring to read how a few individuals, driven by curiosity and faced with strong opposition and criticism, eventually had such a huge impact on science. Yet again, this history illustrates how major breakthroughs, with enormous potential applications and

financial reward, often flow from very basic research. The book makes the point well that good research is driven by new ideas and not by technology. In reality, the technologies available are often inadequate for testing new hypotheses accurately. Thus, it remains puzzling that the private sector remains so reluctant to fund basic research. How many of those who today are reaping the benefits of biotechnology would have been prepared to support this early research financially, especially when controversy reigned in the field? There is probably a lesson to be learned from the fact that Monsanto initiated research on plant genetic engineering in the early 1970s, a decade before the phenomenon was reliably demonstrated.

Plant genetic engineering is a highly controversial topic, and will probably long remain so. Whether the technology will deliver its promise of a better life for all, only time will tell. With the negative impacts of the first Green Revolution still fresh in our minds, many are concerned that the new technology will have an equally negative impact on the environment, and increase the gap between rich and poor still further. Most of the advances in biotechnology are in the hands of major companies in the developed world, and the question is rightly asked whether those most in need will ever be able to afford the technology. It is a pity that the book fails to mention the several instances where it has already significantly improved the living standards of small-scale growers, and the broader community, in the developing world.

Often ignored is the very important role of plant genetic engineering in advancing our understanding of plant metabolism and plant defence mechanisms. As Lurquin points out, many more basic questions regarding collateral gene damage during integration of the new gene into the genome and the control of gene expression must be answered before the technology's potential can be fully realized.

Lurquin's book is the first to describe accurately the history of plant genetic engineering. For students labouring at the bench and getting frustrated at the lack of reproducibility of their experiments, reading this work will provide reassurance. Even scientists who are no longer at the bench, and are now mostly preoccupied with administrative and teaching duties, will find it an important reminder that research is a demanding task, with much disappointment and controversy, and few successes. ■

F. C. Botha is at the Institute for Plant Biotechnology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa.

More on plant biotechnology

Lords of the harvest: Biotech, Big Money and the Future of Food

by Daniel Charles
Perseus Publishing, \$27

Science in culture

Seemingly scientific Oliver Marsden's abstract paintings.

Martin Kemp

There is an increasing range of imagery in non-figurative or wholly abstract art that is unthinkable without the imaging techniques of recent science. And yet the works themselves tend not to illustrate science, and do not even draw specifically on one kind of scientific image. The beguiling and technically impeccable paintings by the young British artist Oliver Marsden are spectacular cases in point.

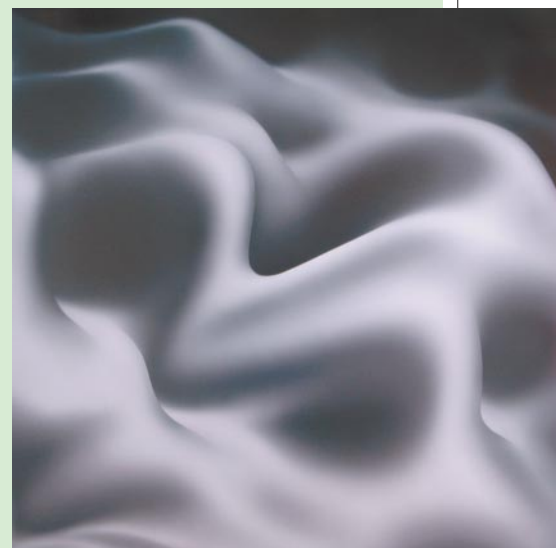
A graduate of the Edinburgh College of Art, Marsden is rapidly marking out a distinctive territory for himself. At the tender age of 28, he has already exhibited internationally, including shows at the Spencer Brownstone Gallery in New York. At first sight, his enticing images, on canvases generally more than a metre square, look as if they are taken from science: from cellular and microbiological formations in the earlier works, or from the undulating surfaces of materials viewed in electron microscopes in recent paintings. But they have acquired their scientific 'look' through Marsden's absorption of the vocabulary of those types of images, rather than because they depict specific substances or phenomena.

The old masters selectively remade specific kinds of natural effect rather than imitating what was in front of them (as they obviously had to do when painting a *Crucifixion* or *Rape of Europa*). Similarly, Marsden uses his understanding of the nature of visual effects in scientific imaging to create forms that speak of, and transform, the visual repertoires of contemporary science.

In particular, his paradoxical forms — which appear real and suggestively solid but tend towards physical impossibility and are ultimately nebulous — are in keeping with the problems that indeterminacy of position and state cause for representation in modern physics.

Marsden's latest exhibition, "Waveform 1 2001", is characteristic of recent work in that it has its origin in manipulated three-dimensional shapes on a computer screen. The results are then photographed as 'sketches' for the paintings. Using a combination of conventional brushes and airbrush sprays on immaculately prepared surfaces, the stunning finish of the images results from an interplay between meticulous control and serendipitous process. Marsden relies on what he calls "balanced chaos".

That his pictures are actually painted and not computer-generated matters greatly to Marsden. Their materiality and the traditional connotations of paint on canvas are integral to their effect and to their dialogue with science. Marsden is also aware that the technique of a hand-made artefact evokes the spectator's awe in a way that computer art still tends not to do.



WaveForm 1 plays tricks with visual grammar.

WaveForm 1 looks as if it is the depiction of something tangible. Yet the sharp, contrasted contours of some forms, which lead us to expect that we are dealing with hard, reflective surfaces, melt into the soft convexities and concavities of particulate clouds. Like the 'pictures' generated by a scanning tunnelling electron microscope, his images obey some of the grammar of things seen within our normal visual compass. But they fail to deliver the full range of internally consistent information about the interplay of light and shade and colour and texture to which we have become accustomed in naturalistic pictures, no less than in nature itself.

Marsden is a keen student of the writings of the nuclear physicist David Bohm, whose notion of 'implicate order' has proved particularly suggestive for artists and non-scientists. Bohm's intuition that there may be a level of order that is inevitably inaccessible to our means of scrutiny is suggestively invoked by the fluidity and ambiguities of Marsden's visual conundrums. The artist brings the time-honoured alchemy of pigment on a flat surface into dialogue with the most advanced techniques of imaging in the physical sciences. It is specifically in the tension between the hand-made and the instrumentally generated that an important facet of the fascination of his paintings lies. ■

Martin Kemp is in the Department of the History of Art, University of Oxford, Littlegate House, St Ebbes, Oxford OX1 1PT, UK. Oliver Marsden's "Waveform 1 2001" is on show at the Blue Gallery, 28/29 Great Sutton Street, London EC1, UK, until 1 December 2001.

Visualizations: The Nature Book of Art and Science is a collection of essays edited by Martin Kemp (published by Oxford University Press and the University of California Press; £20, \$35).

Where might it lead?

Scientists' professional aims should benefit from embracing the 'what if' mentality of science fiction.

Gregory Benford

In my experience, few working scientists read fiction. Typically, they usually read science fiction when young, and then become too busy.

Yet some disparage speculative ideas as "just science fiction", implying that such ideas lack the informing constraints that science can impose. Perhaps this stems from a professional culture that honours strict reporting of observations, conforming to accepted views, and seldom speaking of social impacts.

Speculation on where science might lead us can lurch into tabloid territory and the worst aspects of media fare, UFOs and all. Mega-movies exploit science's spectacles while ignoring its rigour. Technical plausibility and the niceties of proper procedure are roughly shoved aside by drama's drive. Hollywood seldom shows science's subtle pleasures and small victories.

Yet science could better convey much of its mystery and excitement if it paid attention to how writers view it. Ideas usually begin in literature and migrate to the visual media. To alter how science is perceived, we would do well to pay attention to the storytellers and dreamers who are our bards.

Science fiction is our biggest mirror, reflecting the literary imagery of science. Since Mary Shelley's *Frankenstein*, the prospects that science opens have played a significant role in our culture.

We should realize that such cultural currents run both ways. Often science-fiction writers get their ideas from science, whereas the dreams the professionals had when young can fuel their scientific efforts.

The Russian rocketry visionary Konstantin Tsiolkovsky wrote: "The first seeds of [my] ideas were sown by that great, fantastic author, Jules Verne; he directed my thought along certain channels, then came a desire, and after that, the work of the mind." The helicopter pioneer Igor Sikorsky, a compatriot of Tsiolkovsky, became interested in the concept by reading Verne's *Robur the Conqueror*, and cave explorer Norman Casteret was inspired by Verne's *Journey to the Centre of the Earth*. Talents as varied as inventors Guglielmo Marconi and Santos Dumont, explorers Admiral Byrd and Yuri Gagarin, and engineers Simon Lake and Lucius Beebe all paid tribute to Verne's inspiration.

Later science fiction proved to be no less powerful. Leo Szilard began work on

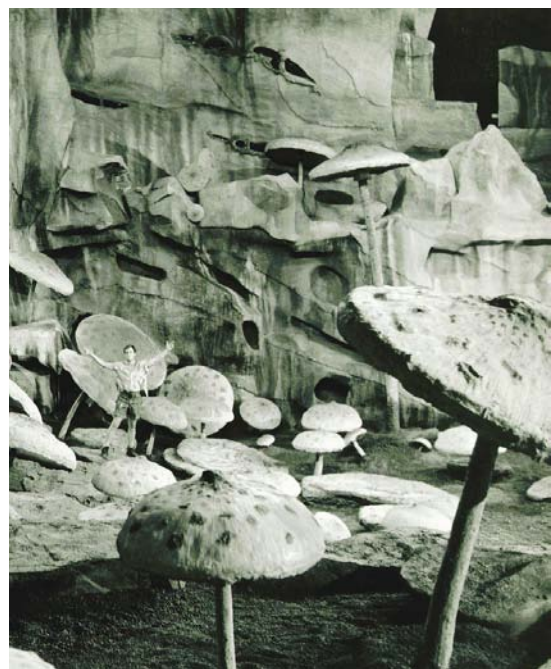
stimulated nuclear fission in 1932 after reading H. G. Wells's *The World Set Free* (1912). Werner von Braun was so addicted to science fiction that he had a subscription to the American science-fiction magazine *Astounding* sent through Sweden and on to him by diplomatic pouch into wartime Germany.

Many inventions first emerged as glints in the eyes of science-fiction authors. Karel Capek's 1921 play *R. U. R.* introduced the word 'robot', and although his creations were organically grown, the term soon came to describe humanoid machines, which populated stories expressing the anxieties of the machine age. Pulp stories of the 1920s anticipated lasers, and vast computers figured in utopian works. In the early 1940s, Robert Heinlein depicted plausible waterbeds and remote-controlled hands in *Waldo*, although his most notable anticipation was the strategic impasse caused by nuclear weapons. Half a dozen stories predicted that the moon landing would be seen on television. As Isaac Asimov once commented: "No one can say that science-fiction writers and readers put a man on the moon all by themselves, but they created a climate of opinion in which the goal became acceptable." More generally, they smoothed the way for science, in its implications and inventions alike.

Editor and critic David Hartwell remarked that "science fiction did not aspire to take over literature, but reality". It "outlasted all other counterculture or outsider literary movements of the century", and now seems to be a less diagrammatic art. Its authors see themselves more as conceptual gardeners, planting seeds for fruitful growth, rather than engineers drawing up blueprints for eternal, grey social machines.

Even now, thinking about ideas, inventions and their social consequences is virtually unknown in mainstream literature. Written science fiction thrashes through

Since Mary Shelley's *Frankenstein*, the prospects that science opens have played a significant role in our culture.



Mushrooming ideas: Jules Verne's novels, such as *Journey to the Centre of the Earth*, inspired many pioneering scientists, explorers and engineers.

thoughtful speculations long before they arrive in common culture.

Alas, few journalists or commentators mine this seam of imaginative experimentation. When Dolly the sheep made cloning a media event, the scientists involved did not notice that science fiction had discussed the issue extensively decades before. Why? Because cloning pokes into the murky realms of our own identity, a personal ground that is often too uncomfortable to discuss until scientists are forced to by events. We prefer abstractions. Both humanists and scientists should see that their professional styles could benefit from more cultural currents, not fewer.

Science captures our abstract wisdom, telling us that we are primates following complex genetic instructions. Writers know that such abstract knowledge does not shelter us from the hammering, immediate moment. Tensions in the lives of scientists, as they move from their day jobs of heady arabesques into their after-hours domestic swarm, are little remarked upon in literature. This largely untravelled territory reminds us that David Hume instructed us to "be a philosopher, but amidst all your philosophy be still a man".

Fiction brings us down to the intensely personal ground of the heart. "We are the music, while the music lasts," as T. S. Eliot put it. But science should also shed its illuminations upon our fevered passions, and should be better seen through fiction's magnifying lens.

Gregory Benford is in the Department of Physics, University of California, Irvine, USA. He is the author of *Timescape*, among other novels.

KOBAL COLLECTION/20TH CENTURY FOX

Parasites go the full monty

Patrick J. Keeling

Energy metabolism is an essential function of life. Yet a resourceful parasite with a minimalist genome has discarded much of its metabolism, developing a unique alternative in the process.

Parasites are well known for stripping themselves down to the bare essentials, but how far will they go? Microsporidia are one group of single-celled parasites that have taken it to extremes and gone the 'full monty'. On page 450 of this issue¹, Katinka and colleagues expose the shocking extent of this striptease in the complete genome sequence of the microsporidian *Encephalitozoon cuniculi*, which parasitizes a range of mammals, including humans. This genome is a mere shadow of those found in other eukaryotes (organisms with nucleated cells), as it consists of only 2.9 million base pairs — less than 0.1% the size of the human genome. This is even smaller than the genomes of many bacteria. Nevertheless, the genome is a treasure trove of information on the powerful reductive forces that shape these unusual parasites.

By any criterion, microsporidia are odd — they can survive only by living inside other cells (Fig. 1), and their method of invading a host cell is truly startling. Microsporidian spores contain a long coiled tube that can be blasted from the spore by a process akin to turning a garden hose inside-out. If this projectile hits a nearby cell, the contents of the spore are forced through the tube and into the cell². This combination of harpoon and hypodermic needle, powered by incredible physical force and requiring seemingly impossible cellular gymnastics, is one of the most sophisticated infection mechanisms in biology.

Yet apart from the structures needed to drive this remarkable machine, and a nucleus to house their DNA, microsporidian spores are pretty spartan. It seems that while they were honing their ability to parasitize other cells, they also abandoned many of the common eukaryotic cell features that biologists take for granted. The *E. cuniculi* genome sequence¹ reflects this reduction at many levels. Missing are genes for a variety of cellular processes and for many metabolic pathways. Moreover, the remaining genes are tightly packed, with little 'junk' DNA between them. Most surprisingly, though, the genes themselves are noticeably shorter than their counterparts in other organisms.

The severe reductionism of microsporidia, together with their amazing specialization, has nearly erased the genomic record of their evolutionary history, making it difficult to determine how they originated. Their simplicity, especially the lack of mitochondria

(the powerhouses that convert sugar to energy in cells), gave rise to the compelling idea that microsporidia might be a primitive eukaryotic lineage that evolved before the acquisition of mitochondria³. Early evolutionary trees lent support to the notion, but recent molecular data suggest that this view needs revising. First, genes derived from mitochondria have been found in microsporidia, showing that their ancestors were in fact endowed with mitochondria^{4–6}. Second, most of the evidence from evolutionary trees now points to microsporidia being highly evolved fungi, rather than ancient eukaryotes⁷. The genome should end this debate, by providing an enormous number of new genes with which to examine microsporidian origins. Indeed, Katinka and colleagues¹ have already identified several genes that strongly support a fungal origin for these parasites. So, characteristics previously considered primitive could now be seen as recent adaptations.

Nevertheless, the apparent lack of mitochondria in microsporidia is still of great interest. Mitochondria are the focal point of energy metabolism in most nucleated cells, so how do microsporidia generate energy without them? Until now, we had very little idea. But Katinka and colleagues use the genome sequence to reconstruct the parasite's core metabolism, with surprising results.

From other organisms, we know that core energy metabolism can be cobbled together in slightly different ways as mitochondria degenerate or are lost, but all these variations share similar adaptations⁸. Microsporidia, however, have invented a unique system of core energy metabolism. They use parts of typical mitochondrial metabolic pathways, mixed with pathways found in other organisms that lack mitochondria, resulting in a system that is fundamentally different from either^{1,9}. In particular, microsporidia lack pyruvate:ferredoxin oxidoreductase — a key enzyme in other organisms that lack mitochondria, and a common target for anti-parasitic drugs — and use mitochondrial pyruvate dehydrogenase in its place^{1,9}. Once again, we see François Jacob's image of evolution as a tinkerer¹⁰, using bits and pieces of existing machines to build a completely new one — in this case, something as central to life as a new form of energy metabolism.

The sheer number of genes for mitochondrial metabolic enzymes retained in the

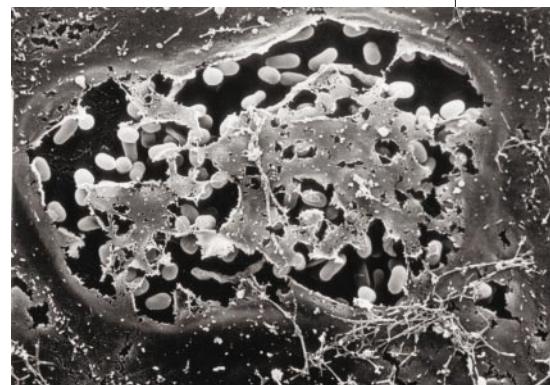


Figure 1 Infection by parasites. A scanning electron micrograph of *Encephalitozoon cuniculi* spores inside a human host cell.

genome is interesting in itself as it raises the question: have microsporidia really shed their mitochondrion, or have we have simply failed to recognize it? This debate deepens with the sequencing of the genome, as it is not absolutely clear from the genes alone whether these enzymes operate in a mitochondrion, or whether they have been relocated to the cytoplasm. There is credible evidence either way^{1,9}, but in this case, the gene sequences only raise the questions. The answers await the localization of the enzymes within the cell.

Although it may seem perverse, parasites are probably the most common and diverse form of life on Earth today. Yet we know little about where they come from, how they end up the way they are, or in some cases even what they are. The genome sequence of *E. cuniculi* is an important step towards answering these questions for microsporidia. Before long, genome sequences from *Plasmodium*, *Cryptosporidium*, *Toxoplasma*, *Entamoeba*, *Trypanosoma*, *Giardia* and others will usher in an age of comparative parasite genomics. This is an exciting time for parasitology, and if *E. cuniculi* is an honest herald, many 'rules' are about to be broken. ■

Patrick J. Keeling is at the Canadian Institute for

- Advanced Research, Department of Botany, University of British Columbia, 3529–6270 University Boulevard, Vancouver, British Columbia, V6T 1Z4, Canada.
e-mail: pkeeling@interchange.ubc.ca
1. Katinka, M. D. et al. *Nature* **414**, 450–453 (2001).
 2. Vávra, J. & Larsson, J. I. R. in *The Microsporidia and Microsporidiosis* (eds Wittner, M. & Weiss, L. M.) 7–84 (ASM, Washington, 1999).
 3. Cavalier-Smith, T. *Nature* **326**, 332–333 (1987).

BRETSILAV KOUDELA

4. Germot, A., Philippe, H. & Le Guyader, H. *Mol. Biochem. Parasitol.* **87**, 159–168 (1997).
5. Hirt, R. P., Healy, B., Vossbrinck, C. R., Canning, E. U. & Embley, T. M. *Curr. Biol.* **7**, 995–998 (1997).
6. Peyretailade, E. *et al.* *Mol. Biol. Evol.* **15**, 683–689 (1998).
7. Keeling, P. J. & McFadden, G. I. *Trends Microbiol.* **6**, 19–23 (1998).

8. Müller M. in *Evolutionary Relationships Among Protozoa* (eds Coombs, G. H., Vickerman, K., Sleigh, M. A. & Warren, A.) 109–132 (Chapman & Hall, London, 1998).
9. Fast, N. M. & Keeling, P. J. *Mol. Biochem. Parasitol.* **117**, 201–209 (2001).
10. Jacob, F. *Science* **196**, 1161–1166 (1977).

Cosmology

The matter with density

Trevor Ponman

Theories of how the Universe developed structure assume a similar distribution of matter in small and large structures. But observations of gas densities in galaxy clusters suggest that this is not the case.

Most galaxies, like our own Milky Way, are part of a small group or larger cluster of galaxies. Ordinary baryonic matter accounts for only 10–20% of the mass of such clusters, the rest being made up of mysterious, and as yet unidentified, 'dark matter'. Most of the baryonic matter is intergalactic gas at a temperature of millions of degrees kelvin, with the galaxies themselves accounting for only around 5% of the total mass. Cosmologists predict that the distribution of matter within large clusters should look very similar to that in small ones — the two being related by a simple spatial scaling. But it has been clear for some time that baryonic matter, at least, does not behave like this: the gas density in the inner regions of small clusters is systematically lower than that in large ones. On page 425 of this issue, Voit and Bryan¹ offer an explanation for these observations, which follows with a certain inevitability from the relationship between the entropy of gas and the time it takes to cool.

Modern cosmology says that structure develops hierarchically. Observations of the cosmic microwave background^{2,3} — the relic radiation left over from the Big Bang — reveal the imprint of small fluctuations in matter density from when the Universe was young. According to theorists, these small fluctuations grow steadily with time, as gravity pulls more mass into regions that are already denser than average. Eventually, the densest regions collapse to form stable structures known as haloes, which are dominated by dark matter, and, as the gas within them cools, star formation results, establishing a galaxy at the centre of each dark-matter halo. When small dark-matter haloes coalesce into larger ones, the galaxies they contain may either merge to form larger galaxies or retain their identity, producing a group of galaxies within a single dark halo. Further mergers of these groups create rich galaxy clusters containing many hundreds of galaxies — these are the largest stable structures found in the Universe today.

One consequence of this hierarchical build-up of structure is 'self-similarity', which says that large dark-matter haloes

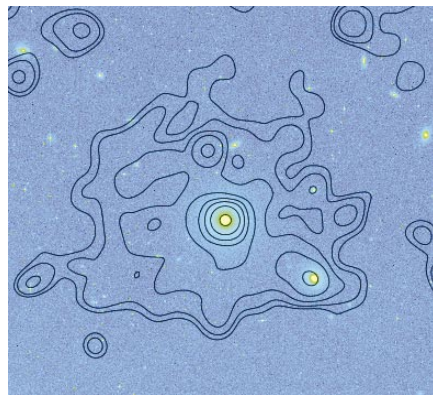


Figure 1 Hot gas in galaxy groups. The contours of X-ray emission recorded by the ROSAT X-ray Observatory are superimposed on an optical image of a galaxy group. Notice the presence of a large central galaxy. The 10-million-degree gas that is responsible for the X-ray emission contains more mass than all the galaxies in the group combined, but its density is lower than expected. Voit and Bryan¹ offer a new explanation of this discrepancy based on the rate at which the gas cools.

are essentially scaled-up versions of small ones. Simple cosmological simulations that include both baryonic and dark matter also predict self-similarity for the distribution of gas within these haloes. Hot, dense gas at 10^6 – 10^8 K emits X-rays, and if self-similarity is true for the gas, then we can expect the X-rays emitted by galaxy clusters (Fig. 1) to have certain scaling properties. For example, larger clusters contain more gas, and have higher characteristic temperatures, and their power output (luminosity, L) in X-rays should scale as the square of the temperature, T . However, it has been clear for many years that this is not the case. The actual relationship is roughly $L \propto T^3$ for large clusters⁴ and is steeper still for small galaxy groups⁵: they are far less luminous than expected. Data from the ROSAT X-ray observatory⁶ revealed the reason for this: the density of gas in the inner regions of small galaxy groups is systematically lower than that in large clusters, contrary to the predictions of self-similarity.

What might explain this discrepancy? Astronomers find it helpful to think in terms of the characteristic entropy of the gas, because this remains unchanged during any adiabatic compression (causing heating) or expansion (causing cooling). Here, entropy is essentially $T/\rho^{2/3}$, where T and ρ are the temperature and density of the gas. The self-similarity model predicts that gas densities should be similar in groups and clusters, so the entropy of the gas should simply scale with T , making it lower in groups, which are cooler. But the entropy of the gas⁶ found in groups is higher than expected, because at a given temperature its density is lower.

These observations support the idea that there is a critical 'floor' value below which the entropy cannot drop. In clusters, most of the gas would naturally lie above this floor value anyway. But in groups, where the entropy is lower as a result of the low temperature, this floor prevents the gas from reaching the high densities required for bright X-ray emission. It has been suggested^{7,8} that this critical value might be set by preheating of the gas — by the energy released in supernova explosions or from quasars — before it was gathered into clusters, but the amount of energy required for this is uncomfortably large.

Voit and Bryan¹ take a different line. They suggest that cooling of the gas, rather than heating, sets the critical value. This is because gas with low entropy cools more quickly than high-entropy gas, so the critical entropy value is set by the gas whose cooling time is equal to the age of the Universe. The authors calculate this critical entropy value, and find that it agrees remarkably well with the observed floor value.

Gas with entropy below the floor value will suffer one of two fates: it either cools and disappears from the hot-gas phase to form stars, or it can be reheated by star formation to above the critical value. Either way, the result is to remove the low-entropy gas that would otherwise dominate the inner regions of clusters and make their X-ray emission more luminous. By simply removing such gas from their models, Voit and Bryan can reproduce the observed luminosity–temperature relation for groups and clusters. To prevent much of the gas in the Universe from cooling within small structures, some form of heating is still required. But the energy needed is likely to be lower than in earlier theories, because the heating is naturally targeted at the low-entropy gas, as this is where star formation, and therefore supernova explosions, occur.

These important insights are not the end of the story. First, we do not yet know the balance between cooling and reheating. In order to establish whether most of the low-entropy gas has been removed by cooling (which removes it from the hot-gas phase altogether) or by reheating (which does not), we need to measure the total gas content of

galaxy groups. This is not easy, because their X-ray emission is weaker than that of large clusters. Second, reheating from star formation may not be the whole story. Bright galaxies are found at the centre of many clusters, and it is now widely believed that these harbour central black holes. Gas cooling onto these galaxies may serve to fuel the black hole, and it is possible that this, rather than supernova explosions, serves to reheat the low-entropy gas^{9,10}.

Solutions to these problems may soon be provided by the new generation of X-ray observatories now in operation in space. XMM-Newton is sensitive enough to measure the amount of gas in galaxy groups, whereas Chandra has the spatial resolution to study the interaction between central active galaxies and the gas surrounding

them. These observational advances can be expected to stimulate further theoretical developments over the next few years. ■

Trevor Ponman is at the School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK.
e-mail: tjp@star.sr.bham.ac.uk

1. Voit, G. M. & Bryan, G. L. *Nature* **414**, 425–427 (2001).
2. de Bernardis, P. *et al.* *Nature* **404**, 955–959 (2000).
3. Hanany, S. *et al.* *Astrophys. J.* **545**, L5–L9 (2000).
4. Edge, A. C. & Stewart, G. C. *Mon. Not. R. Astron. Soc.* **252**, 414–441 (1991).
5. Helsdon, S. H. & Ponman, T. J. *Mon. Not. R. Astron. Soc.* **319**, 933–938 (2000).
6. Ponman, T. J., Cannon, D. B. & Navarro, J. F. *Nature* **397**, 135–137 (1999).
7. Kaiser, N. *Astrophys. J.* **383**, 104–111 (1991).
8. Evrard, A. E. & Henry, J. P. *Astrophys. J.* **383**, 95–103 (1991).
9. Valageas, P. & Silk, J. *Astron. Astrophys.* **350**, 725–742 (1999).
10. Wu, K. K. S., Fabian, A. C. & Nulsen, P. E. J. *Mon. Not. R. Astron. Soc.* **318**, 889–912 (2000).

Evolution

Tides of tolerance

Karl Sigmund and Martin A. Nowak

Humans, and many other species, have a tendency to cooperate and help each other. But how does such behaviour evolve? Some new computer simulations provide a plausible answer.

When Charles Darwin¹ published his theory of evolution in 1859, he knew that cooperation and altruistic behaviour present something of a problem for a concept that is based on competition and the struggle for existence. He did, however, anticipate a solution that was provided by William Hamilton² more than a century later: cooperation can emerge as a result of 'kin selection' in cases where interacting individuals are genetically related. On page 441 of this issue, Riolo and colleagues³ discuss a new model for the evolution of cooperation, in which individuals help others that are, in some way, like themselves.

This is not the first time that the idea of 'like helping like' has been suggested as a route to the evolution of cooperation. Twenty-five years ago, Dawkins⁴ introduced the 'green beard effect' as a thought experiment in sociobiology. Consider a gene that confers on its bearer not only a green beard (or any other distinctive label), but also the instinct to provide assistance to all other owners of a green beard. Individuals with such a gene would effectively form a self-serving clique, and so the gene would spread within the population.

Today, the green beard is a cherished icon of the 'selfish gene' view of natural selection, and similar effects have actually been found in nature⁵. But it still takes some effort to accept the idea of a gene producing, simultaneously, a signal and a predilection for the signaller. Such a double-effect gene seems contrived. Riolo *et al.*³ discuss a more plausible model for the evolution of cooperation:

individuals just need to like their like, something that most of us can relate to.

In traditional models of how cooperation can emerge⁶, the evolutionary development of a fictitious population of 'agents' is simulated, on a computer, over many generations, with pairs of individuals meeting randomly as potential givers and receivers of help. Giving help entails some cost to the donor, and getting help provides a larger benefit to the recipient. Cost and benefit are measured in reproductive success, and offspring are supposed to inherit the parent's behaviour, unless mutations occur. Because 'selfish' individuals, who refrain from helping, incur no costs, they spread, and after some generations cooperation will be eliminated. But if individuals can guess whether recipients are likely to give assistance in their turn, they can channel help towards those who help, and discriminate against exploiters. In this way cooperation, based on reciprocation, can emerge⁶.

Riolo *et al.*'s model³ follows a similar pattern, but with some crucial differences. Each individual has a trait (or 'tag', as Riolo *et al.* call it), and is endowed with a tolerance level. Donors refuse to offer help if the trait of the recipient differs from their own by more than the donor's tolerance level. In other words, zero tolerance means that an individual helps only those who are exactly like it; maximum tolerance means helping everyone. Both the trait and the tolerance level are inherited by the offspring, with some mutations occasionally introducing new variations into the population.

The basic outcome of the computer simulations³ is that a substantial degree of cooperation is established. Essentially, a 'dominant cluster' emerges, consisting of players sufficiently similar to help each other. This occurs even in the absence of repeated interactions and reputation effects — that is, without direct or indirect reciprocation. All that is needed is some recognition of what is 'similar', an ability that is widespread among animals⁷ (odours or visual cues can provide the required information). So, the mechanism that leads to cooperation is a form of kin selection — either classical (if traits are inherited genetically) or social (if they are inherited culturally, like a dress code).

One attractive feature of the new simulations is the evolution of tolerance — the recognition mechanism that discriminates 'us' from 'them'. This tolerance does not freeze at some fixed value. Cyclically, it slowly increases over time, and then sharply declines. This drop occurs when the dominant cluster is dissolved from within as a result of mutation, by new individuals whose traits lie in the range of the dominant cluster but whose tolerance is considerably reduced. These newcomers are helped by all the residents of the established cluster but themselves help just a few, so they bear fewer costs than the established residents. A wave of intolerance then sweeps through the population, and in its wake a reduction in overall cooperation. But once a new dominant cluster is established, cooperation resumes at its former level and tolerance starts spreading again. The slow upward drift of tolerance seems to be due to a combination of mutational pressure and kin selection. It will be important in the future to explore the robustness of this phenomenon.

These oscillations of tolerance levels are striking, and bring to mind many historical instances. We are witnessing a wave of social and religious intolerance right now. It would be foolish, of course, to reduce the complexities of political life to the vagaries of a virtual population. Yet these computer simulations do capture the imagination, and may well lead to a cottage industry of follow-up investigations, just like Axelrod's famous computer tournaments based on the 'prisoners' dilemma' game⁸.

The new scenario applies to both genetic and cultural tags⁹. Part of its appeal is its obvious link to reality — school ties, club memberships, tribal customs or religious creeds are all tags that induce cooperation. Some of these tags are easy to fake and might invite exploitation. Language, on the other hand, could be a reliable tag that is hard to fake; hiding one's accent in a foreign language is nearly impossible.

Furthermore, tags can help to encourage the usual suspects behind cooperation among unrelated individuals: direct and indirect reciprocation (whereby recipients

of help return the help, either to the donor or to a third party). In other words, tags bolster the emergence of cooperation in repeated interactions, and might even promote long-term pairings based on similarity¹⁰. Indeed, to find out how much one has in common is one of the first delights of falling in love and contemplating a lifelong partnership.

But tags can also present major obstacles in overcoming segregation. Although the simulations by Riolo *et al.*³ do not produce dominant clusters that split into rival tribes, any territorial distribution would favour such 'speciation'. Tags would then act as self-enforcing stereotypes, making it hard for tolerance to cross the divide. We know that discrimination is needed to sustain cooperation in the face of exploiters. But tag-based intolerance could turn discrimination away from the 'bad guys' and raise senseless antagonism. ■

Surface chemistry

Catalysis frozen in time

J. K. Nørskov

Modern microscopes are not just for imaging. In the right hands they can be used to follow and control catalytic reactions on a metal surface — one atom at a time.

Solid surfaces act as catalysts for a large number of reactions: it is estimated that 20–30% of the gross national product in developed countries is dependent one way or another on this sort of catalysis¹. The surface acts by adsorbing the reactants and encouraging them to react until the products leave the surface. Yet despite its importance, many aspects of catalysis by solid surfaces are not understood. The use of powerful tools is now helping to change that. Writing in *Physical Review Letters*, Hahn and Ho² show that they can manipulate individual atoms and molecules adsorbed on a metal surface to induce a catalytic reaction. By using this technique at low temperatures they can control the speed of the reaction, so they can follow important steps as they happen, atom by atom.

The catalytic oxidation of carbon monoxide (CO) is one of the simplest catalytic reactions, and is an important part of the reactions that take place in the catalytic converter in your car. In this context, CO is transformed into carbon dioxide (CO₂) by reacting with oxygen or nitrous oxide (NO) on the surface of a platinum, palladium or platinum–rhenium catalyst³. The reaction involves five steps, which are summarized in Fig. 1. At the low temperatures (13–45 K) used by Hahn and Ho, the two first steps (Fig. 1a) can occur on a silver surface, but the rest of the reaction requires higher temperatures to proceed spontaneously.

At these low temperatures the authors use

Karl Sigmund is at the Institut für Mathematik, Universität Wien, Strudlhofgasse 4, A-1090 Vienna, and the International Institute for Applied Systems Analysis, Laxenburg, A-2361 Austria.

e-mail: karl.sigmund@univie.ac.at

Martin A. Nowak is at the Institute for Advanced Study, Princeton, New Jersey 08540, USA.

e-mail: nowak@ias.edu

1. Darwin, C. *On the Origin of Species by Means of Natural Selection* (John Murray, London, 1859).
2. Hamilton, W. D. J. *Theor. Biol.* **7**, 1–16 (1964).
3. Riolo, R., Cohen, M. D. & Axelrod, R. *Nature* **414**, 441–443 (2001).
4. Dawkins, R. *The Selfish Gene* (Oxford Univ. Press, New York, 1976).
5. Keller, L. & Ross, K. *Nature* **394**, 573–575 (1998).
6. Nowak, M. A. & Sigmund, K. *Nature* **393**, 573–577 (1998).
7. Pfennig, D. W. & Sherman, P. W. *Sci. Am.* **272**, 98–103 (1995).
8. Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
9. Boyd, R. & Richerson, P. *Culture and the Evolutionary Process* (Univ. Chicago Press, 1985).
10. Riolo, R. L. in *Proceedings of the 7th International Conference on Genetic Algorithms* (ed. Bäck, T.) 378–385 (Morgan Kaufmann, San Francisco, 1997).

a scanning tunnelling microscope (STM) to make the reaction happen and to image and manipulate the atoms and molecules on the surface (Fig. 1). An STM works by applying a voltage to a sharp metallic 'tip', which is scanned across the surface, and then recording the flow of electrons that tunnel between the surface and the tip. In practice, the flow of electrons is kept constant by adjusting the distance between the tip and the surface. When the tip is scanned over the entire surface, a record of its height at each point provides a detailed map of the surface at atomic resolution⁴. The STM can also be used to physically modify the surface: the electrons transferred from or to the tip can start a chemical reaction, or make atoms or molecules move on the surface^{5,6}. Pushing or pulling an individual atom or molecule with the tip can also make it move.

Hahn and Ho use the STM to dissociate the oxygen molecules (Fig. 1b), to move individual CO molecules close to the oxygen atoms on the surface, and to induce the final reaction and desorption of the product (Fig. 1c, d). Along the way, they image the intermediates on the surface — they observe, for instance, a complex consisting of two oxygen atoms close to each other and to a CO molecule. They also used the STM in its 'inelastic electron tunnelling spectroscopy' mode to monitor the vibrational behaviour of CO as it is nudged closer and closer to the two oxygen atoms. The formation of the

O–CO–O complex is confirmed by observing a change in the vibrational frequency of the CO. Measuring the spatial distribution of the vibrational intensity of CO within the complex provides information about the structure of the reactants. In a separate set of experiments they induced the reaction by first transferring the CO molecule to the tip (by using a voltage pulse), moving the tip into position over an adsorbed oxygen atom, and applying a new voltage pulse with the opposite sign to transfer the molecule back to the surface and to kick start the reaction.

These experiments provide atomic-scale details of a chemical reaction occurring at a surface. Intermediates that would not have a measurable lifetime at higher temperatures, and so cannot be observed during a thermal reaction, can be viewed directly. The clever part is to work at temperatures low enough for the intermediates to be frozen in time and to use electron injection rather than thermal excitations to make the reaction proceed.

The work of Hahn and Ho shows that it is possible to induce reactions that won't proceed thermally by using tunnelling electrons to activate the reaction. Using an STM tip to induce chemical reactions at a catalytic surface is not an efficient way of producing large amounts of chemicals, but it is analogous to the way natural catalysts (enzymes) manage

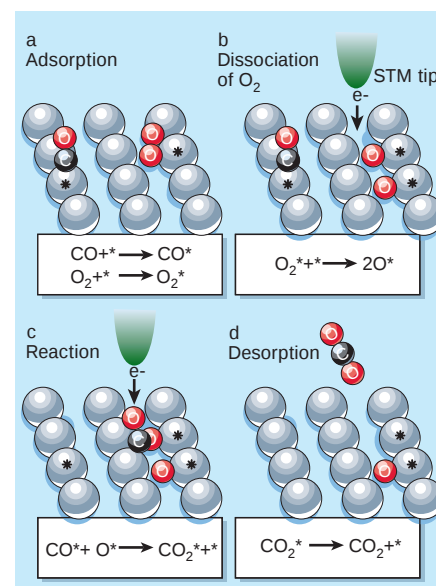


Figure 1 Surface chemistry in action. The formation of carbon dioxide (CO₂) by catalytic oxidation of carbon monoxide (CO) on a metal surface is an important reaction in air purification, emission control and chemical sensing. In their experiment², Hahn and Ho show that CO and oxygen atoms (O) adsorbed on a silver surface can be transformed into CO₂ with the help of an STM tip, which transfers electrons to the reactants. With this technique the reaction can proceed at temperatures so low that every step of the reaction (a–d) can be watched closely. (An asterisk denotes a site on the surface that can form a bond to an adsorbed atom or molecule.)

to do difficult reactions at room temperature. The enzyme nitrogenase, for instance, can activate the very stable N–N bond in N_2 at room temperature^{7,8}, whereas an industrial iron-based catalyst⁹ needs a temperature of about 400 °C. Part of the reason for this difference may be related to the way in which electrons are injected into the active site in the enzyme using energy released during ATP hydrolysis^{7,10}. Nature may already be using the same general principle as Hahn and Ho to induce reactions at temperatures where they would not proceed thermally. ■

J. K. Nørskov is at the Centre for Atomic-Scale

Materials Physics, Department of Physics, Technical University of Denmark, DK-2800 Lyngby, Denmark.

e-mail: norskov@fysik.dtu.dk

1. Maxwell, I. *Stud. Surf. Sci. Catal.* **101**, 1–9 (1996).
2. Hahn, J. R. & Ho, W. *Phys. Rev. Lett.* **87**, 166102 (2001).
3. Nieuwenhuys, B. E. *Surf. Rev. Lett.* **50**, 1869–1888 (1996).
4. Besenbacher, F. *Rep. Prog. Phys.* **59**, 1737–1802 (1996).
5. Lee, H. J. & Ho, W. *Science* **286**, 1737–1802 (1999).
6. Hla, S. W., Bartels, L., Meyer, G. & Rieder, K. H. *Phys. Rev. Lett.* **85**, 2777–2780 (2000).
7. Burgess, B. K. & Lowe, D. J. *Chem. Rev.* **96**, 2983–3012 (1996).
8. Rod, T. H. & Nørskov, J. K. *J. Am. Chem. Soc.* **122**, 12751–12763 (2000).
9. Nielsen, A. *Catal. Rev.* **4**, 1–26 (1997).
10. Rod, T. H., Logadottir, A. & Nørskov, J. K. *J. Chem. Phys.* **112**, 5343–5347 (2000).

Ecology

The guts of seed dispersal

Peter D. Moore

Some plants have evolved delicious fruits to entice animals and birds into dispersing their seeds. But the seeds' dormancy and germinability are affected by which bird they pass through.

Hitching a lift in the gut of a bird is a pretty effective answer to the mobility problem that faces every plant. A bird eats a fruit and flies off, in due course depositing the seeds it has ingested elsewhere, allowing a new plant generation a fresh start in life. But passing through an avian alimentary system can also affect the overall viability of seeds, as well as their dormancy characteristics and germination rate. These conclusions emerge from a series of experiments with birds and berries on the Balearic Islands, conducted by A. Traveset and colleagues, and published in the latest issue of *Functional Ecology*¹. The authors also tackle the question of whether different bird species selectively modify their internal plant passengers, introducing another variable to be taken into account in considering the ecology of vegetation regeneration.

Some birds are seed predators, destroying

and digesting seeds for their nutrient content. But even these can act as dispersal agents if they die with viable seeds in their gizzards, or if some seeds fortuitously evade destruction in the gut. Other birds consume entire fruits and digest only the fleshy parts, allowing the seeds to pass on through the gut. These can be efficient dispersers.

Darwin² speculated about the dispersal of seeds in bird guts, calculating — perhaps optimistically — that a propagule could travel 500 miles in this way before the host evacuated its bowels or died. It has also been recognized that passage through the gut can increase the germinability of seeds³, a quality known to feature in plant ecological strategies⁴. But the mechanisms that affect germination properties within the gut, and the degree to which different bird species vary in their modification of seed viability, have not been studied in any detail.

Traveset and colleagues¹ have taken an experimental approach. They selected two fruit-eating birds (blackbird, *Turdus merula* (Fig. 1) and Sardinian warbler, *Sylvia melanocephala*) and five species of plant from the Mediterranean scrub vegetation of the Balearic Islands, and tested the effects of gut passage on the seeds. At the same time, the fleshy part of control fruits were removed by hand and the seeds checked for their germination potential. Three characteristics were tested: dormancy (time taken from sowing to first germination); germination rate (the proportion germinating within a set time); and germinability (the final proportion of seeds that germinate). Seeds were also examined for changes that resulted from gut passage, such as weight change, or — using electron microscopy — alteration in the thickness or structure of the seed coat.

The first point to emerge from the experiments is that the species of bird that eats the fruit has an influence on the outcome. In the case of the evergreen shrub *Osyris alba*, for example, the germination rate and overall germinability were greatly enhanced by passing through the gut of a blackbird. In bramble, *Rubus ulmifolius*, on the other hand, consumption by the warbler significantly reduced the germination rate of the seeds. Passage through blackbirds also decreased the germination rate, but to a lesser degree. In the case of the madder, *Rubia perigrina*, ingestion by blackbirds both delayed germination and reduced germination rate, whereas the warbler had a lesser effect on the rate. For *Asparagus acutifolius*, blackbirds slightly enhanced the germination rate of seeds, though not significantly, but their consumption of *Asparagus* fruits led to much faster seed germination. The fruits of *Asparagus* and *Osyris* are too big for the Sardinian warbler to swallow, so comparisons could not be made with this bird.

Traveset *et al.* found that seed weights after gut passage were generally reduced when compared with controls, but this was not due to a thinning of the seed coats. No structural changes were detected in the coats, but it is possible that chemical change or scarification — physical scratching rather than thinning — could account for the weight loss. Damage to the seed coat in the gut of birds, particularly in their muscular gizzards, is often invoked as the reason for germination enhancement, but how germination can be delayed as a consequence of gut passage remains unknown. Differing gut retention times among bird species may have an effect. These times are often short — as little as nine minutes for some species⁵ — but can be up to an hour or more in blackbirds⁶. Little is known of the changes that occur in the gut during this period.

Two ecological conclusions emerge from these experiments. First, the germination behaviour of seeds may be modified in various ways as a result of gut passage, by either shortening or lengthening the period of dormancy. Second, different bird species have different effects on a given species of plant. The control of germination time in seeds is known to have repercussions in terms of the overall ability of a plant to grow and reproduce⁷. Variability in germination times can have considerable advantages for a plant, particularly when it has to deal with unpredictable habitats or those prone to catastrophic events⁸. The destruction of a cohort of seedlings by a factor such as drought can be counteracted by a second wave of germination by dormant seeds in the soil. Early germination, on the other hand, may assist the seed in escaping predation or fungal attack, or allow it to take advantage of unusual conditions, such as the early arrival of rain. So a combination of both options —

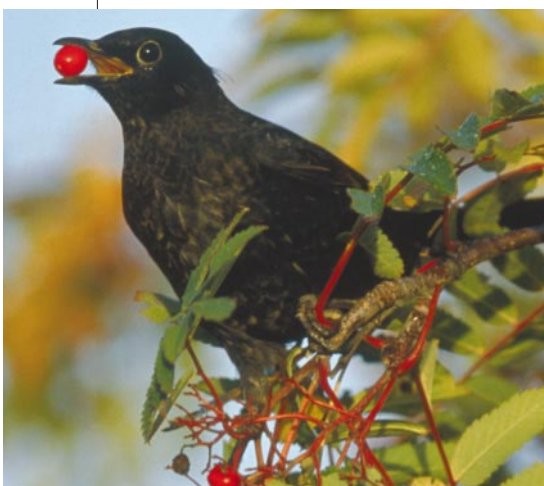


Figure 1 Blackbird (*Turdus merula*) feeding on fruit.

achieved by the fruit passing through different birds — could diversify the germination strategies available to the plant. For a seed travelling by bird, therefore, the long haul is by no means the only benefit available; the choice of transport and the in-flight service could be even more important. ■

Peter D. Moore is in the Division of Life Sciences, King's College, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NN, UK. e-mail: peter.moore@kcl.ac.uk

1. Traveset, A., Riera, N. & Mas, R. E. *Funct. Ecol.* **15**, 669–675 (2001).
2. Darwin, C. *The Origin of Species* 6th edn (Murray, London, 1892).
3. Fenner, M. *Seeds: The Ecology of Regeneration in Plant Communities* (CAB International, Wallingford, Oxon, 1992).
4. Grime, J. P. *Plant Strategies, Vegetation Processes, and Ecosystem Properties* 2nd edn (Wiley, Chichester, 2001).
5. Levey, D. J. in *Frugivores and Seed Dispersal* (eds Estrada, A. & Fleming, T. H.) 147–158 (Junk, Dordrecht, 1986).
6. Barnea, A., Yom-Tov, Y. & Friedman, J. *Funct. Ecol.* **5**, 394–402 (1991).
7. Kelly, M. G. & Levin, D. A. *J. Ecol.* **85**, 755–766 (1997).
8. Baker, H. G. in *The Genetics of Colonising Species* (eds Baker, H. G. & Stebbins, G. L.) 147–168 (Academic, New York, 1965).

Palaeontology

On being vetulicolian

Henry Gee

Some curious fossils from the Cambrian period have been grouped into a new phylum, the Vetulicolia. All of its members are extinct, and their unusual anatomy tempts evolutionary speculation.

Researchers engaged in acts of discovery sometimes have to confront the truly strange and make sense of it. Many of the fossils from the Cambrian Period (about 543–510 million years ago) are cases in point, for they defy comparison with living creatures. In this seeming 'explosion' of life, evolution was toying with many different designs. If all living things share a common ancestor, even the strangest fossil creature must be related to something more intelligible, whether living or long-since dead. Identifying these relatives can be a daunting task.

Such is the problem faced by Shu and colleagues, who on page 419 of this issue¹ describe and discuss several fossil animals from the Chengjiang fauna of southern China. This is a remarkable assemblage of exotic forms, comparable with — and slightly older than — the fossils of the famous Burgess Shale of western Canada. Shu *et al.* describe the exquisite fossils of an unusual, partially segmented creature (see Fig. 1 on page 420) and, in an extended comparison with three other known fossils (Figs 2–5), create a new phylum — a natural group of animals united by a distinctive body plan. They suggest that this phylum, which they call the Vetulicolia, is part of the Deuterostomia, the 'superphyletic' group that includes echinoderms (starfishes, sea urchins and allies), hemichordates (the obscure acorn-worms and pterobranchs) and chordates (vertebrates, including ourselves, as well as tunicates and the amphioxus *Branchiostoma*). The relationships of these various groups are outlined in Fig. 1 overleaf.

Interpretation of any problematic fossil is a risky business. Interpreting presumably adult forms of an unusual fossil as deuterostomes is riskier still, because the defining characters of deuterostomes are conventionally embryological and will not show up in adults. However, molecular work suggests

there is a close alliance between hemichordates, which have structures called pharyngeal slits, and echinoderms, which do not, at least in living forms². This conclusion implies that some features seen in adult chordates and hemichordates — in particular the pharyngeal slits — might be characteristics of deuterostomes in general. Shu and colleagues present evidence that vetulicolians, like hemichordates and chordates, have pharyngeal slits; this finding, among others, aligns them with deuterostomes. If this interpretation is correct, vetulicolians represent a new, primitive deuterostome body plan that could shed light on the long vexed question of vertebrate origins.

The vetulicolian fossils are just a few centimetres long, and are united by several characteristics, including a markedly bipartite body. The anterior half is voluminous and sac-like, with a large opening at the front — regarded as the mouth — and five pairs of lateral openings. Shu and colleagues interpret these features as pharyngeal slits. At least some vetulicolians have a trace of a groove or cleft on the inside surface of the floor of the anterior cavity, which Shu *et al.* interpret as an endostyle. The endostyle — a characteristic feature of chordates — is a gland-rich gutter in the ventral floor of the pharynx. The mucus it produces lubricates the inside of the pharynx, where it traps and concentrates food particles. It also, incidentally, concentrates iodine. The candidate endostyle of vetulicolians is similar in position to those of tunicates and the amphioxus. The filter-feeding larvae of lampreys are the only vertebrates to retain an endostyle. On metamorphosis, the endostyle becomes the adult thyroid gland. Neither echinoderms nor hemichordates are thought to have an endostyle, which appears to be a uniquely chordate feature.

The posterior half of the vetulicolians is divided into seven segments. A structure



100 YEARS AGO

The Leonid Meteors, November, 1901. A telegram to the daily Press through Reuter's agency announces that a considerable number of meteors have been observed in localities where the weather conditions were propitious. Advices from many stations in the United States report more or less brilliant displays of the Leonids as having been seen on Thursday and Friday nights. A steamer from New Orleans reports having seen a great shower near Cape Hatteras early on Friday morning (November 15). The only night on which the sky was at all favourable in London was Thursday, November 14, and on that occasion continual watch was kept by three observers at the Solar Physics Observatory from 11 p.m. to 4 a.m. A few meteors were seen, from twenty to thirty, but nothing in the semblance of a definite shower was presented. Many of the shooting stars seen were very brilliant, but those traced out as being Perseids or Taurids were as numerous as those decidedly radiating from the sickle of Leo, so that probably there was nothing more than is to be seen on any good night for the same interval of time. From *Nature* 21 November 1901.

50 YEARS AGO

In the last century, the idea of a universal and all-pervading æther was popular as a foundation on which to build the theory of electromagnetic phenomena. The situation was profoundly influenced in 1905 by Einstein's discovery of the principle of relativity, leading to the requirement of a four-dimensional formulation of all natural laws. It was soon found that the existence of an æther could not be fitted in with relativity, and since relativity was well established, the æther was abandoned. Physical knowledge has advanced very much since 1905, notably by the arrival of quantum mechanics, and the situation has again changed. If one re-examines the question in the light of present-day knowledge, one finds that the æther is no longer ruled out by relativity, and good reasons can now be advanced for postulating an æther... We can now see that we may very well have an æther, subject to quantum mechanics and conforming to relativity, provided we are willing to consider the perfect vacuum as an idealized state, not attainable in practice... It is no longer a trivial state, but needs elaborate mathematics for its description. P. A. M. Dirac From *Nature* 24 November 1951.

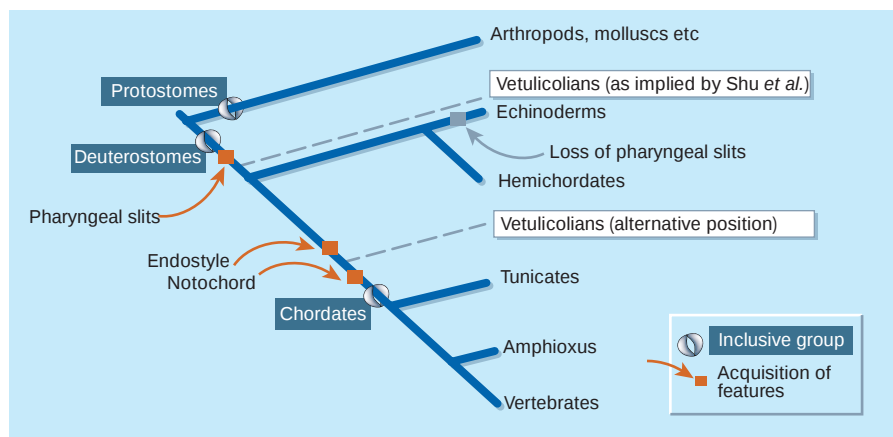


Figure 1 The deuterostomes, and possible places for vetulicolians in the evolutionary scheme of events. Shu and colleagues¹ regard vetulicolians as 'basal' deuterostomes — phylogenetically outside all extant deuterostomes. But they also suggest that the fossils have an endostyle, a structure generally seen as characteristic of a more exclusive deuterostome group, the chordates, and implying that the vetulicolians have a more 'crownward' position in the tree of life. As interpreted by Shu and colleagues, the fossils lack a notochord and so cannot be regarded as chordates. But vetulicolians could be their immediate sister taxon.

interpreted as a gut runs along the entire length of the posterior section, which looks remarkably arthropod-like. Indeed, the authors remark that the entire animal looks much like a small shrimp. But vetulicolians appear to have been limbless, and no arthropod exists that has traded all its limbs for pharyngeal slits.

Where do the vetulicolians fit into deuterostome evolution? The authors' assertion that these creatures were 'basal deuterostomes' — that is, deuterostomes that branched off the main lineage before the echinoderms, hemichordates and chordates became distinct — is properly cautious, given the difficulties inherent in interpreting unusual fossils. To assert that vetulicolians were more or less close to specific deuterostome groups might be asking more of the evidence than it can stand. External commentators, however, like court jesters, have licence to be more reckless.

Looking at the distribution of features in vetulicolians, and mapping them onto a deuterostome phylogeny (Fig. 1), you could make a case that these animals are akin to chordates, rather than deuterostomes in general. This would then make vetulicolians directly relevant to the question of vertebrate ancestry. In short, the latest common ancestor of chordates would have looked rather like a vetulicolian. The only real difference would be that vetulicolians lack a notochord — the axial, stiffening rod that all chordates have at some stage in their life cycle. The amphioxus has a notochord throughout life; tunicates lose it as adults, and in most vertebrates it is supplanted by the vertebral column. Shu and colleagues do not find compelling evidence for a notochord in vetulicolians.

Nonetheless, the vetulicolian body plan is close to what could be regarded as arche-

typal for the most primitive chordates. In the early 1970s, A. S. Romer³ speculated that the vertebrate body was an amalgam of two separate entities: the 'visceral' and the 'somatic'. The visceral animal corresponds with the internal organs, musculature (characteristically smooth) and associated

innervation; the somatic part includes the skeleton, the musculature of the body wall (generally striated), the central nervous system and the sense organs.

Romer used this 'dual-animal' model as a device to explain the course of vertebrate evolution. He argued that primitive chordates, such as tunicates, are all viscera — little more than a sac-like pharynx and gut with the most rudimentary neuronal apparatus. Vertebrates evolved by developing the somatic part, originally as the organ of locomotion in the posterior part of the animal — a beginning can be seen in the locomotory tail of tunicate tadpole larvae. In vertebrates, the somatic part has grown forwards and dorsally, covering and finally encapsulating the visceral part.

Perhaps the most striking feature of the vetulicolians is the strong division of the animal into anterior and posterior halves, just as Romer speculated. To a mind attuned to prospective vertebrate ancestors, it is easy to look at a vetulicolian and see — for example — a tunicate tadpole larva.

Henry Gee is a senior editor at Nature.

e-mail: h.gee@nature.com

1. Shu, D.-G. *et al.* *Nature* **414**, 419–424 (2001).

2. Halanych, K. M. *Mol. Phylogenet. Evol.* **4**, 72–76 (1995).

3. Romer, A. S. *Evol. Biol.* **6**, 121–156 (1972).

Superconductivity

Copper oxides get charged up

Allan Hugh MacDonald

A technique for injecting electrons into the surface layers of materials has now been applied to the most mysterious of superconducting compounds — the copper oxides.

The discovery of superconductivity in copper oxide materials¹ in 1986 led to an explosion of research into these compounds. They became known as high-temperature superconductors because they have negligible resistance to electric current at temperatures up to 160 K, much higher than other materials known at the time. They continue to excite interest because many of their properties cannot be explained by the quantum theory of solids developed over the past 70 years. To many researchers it seems² that something fundamentally new is afoot. On page 434 of this issue Schön *et al.*³ describe a new way of preparing these copper oxide superconductors that makes it easier to probe important aspects of their behaviour.

The superconducting properties of these materials are known to be sensitive to the density of mobile electrons in the CuO₂ planes. In particular, small changes in electron density can transform these materials from an insulating into a superconducting

state, and vice versa. In the past, the electron density has been modified by adding impurities to alter the chemical structure, a process known as 'doping'. However, the impurities inevitably introduce disorder, which can change the properties of the superconductor in a way that has nothing to do with changes in their electron density. Schön *et al.* have developed a system in which an undoped copper oxide material can be made superconducting by injecting charges from a metal contact into the material. This technique makes it possible, for the first time, to investigate how the properties of a given copper oxide superconductor vary with electron density, without the complicating influence of disorder.

The simplest description of electrons in a solid-state crystal is based on an approximation in which each electron moves independently, sensitive only to the average location of other electrons. It turns out that this approximation implies that crystals with odd numbers of electrons per molecular unit

must be metals. This observation, made shortly after the development of quantum mechanics, helped to explain the electronic properties of many solids. But copper oxide compounds with an odd number of electrons per copper atom are insulators. Their charge transport is blocked because the energy cost of moving an electron from one atom to a second, where it experiences repulsion from existing electrons, is too large. Insulators of this type — including the copper oxide material studied by Schön and colleagues — are known as Mott insulators.

Electrons in a Mott insulator are localized on atoms, and their magnetic moments, or 'spins', prefer to have an ordered arrangement in which neighbouring spins point in opposite directions, producing an antiferromagnetic state. Such electrons cannot move independently; a theorist trying to understand the rules of their correlated motion is like an outsider trying to follow the steps of a charming but unfamiliar country dance. Everyone finds their place, but how? The Mott-like behaviour of copper oxide superconductors suggests⁴ that understanding their electronic correlations is essential to understanding their superconductivity.

Superconductivity in copper oxides occurs when the number of electrons per copper atom differs slightly from the integer value that gives rise to a Mott insulator. It can be fractionally larger if it has been doped with atoms that add electrons to the CuO₂ planes (electron doped) or smaller if it is doped with atoms that remove electrons from the CuO₂ planes (hole doped). The superconducting state in 'underdoped' copper oxides — those with relatively low levels of doping — has proved particularly difficult to understand. For example, angle-resolved photoemission experiments^{5,6} have shown that characteristic excitation energies in underdoped copper oxides do not drop to zero at the superconducting transition temperature, as they do in conventional superconductors. Moreover, experiments using neutrons and scanning-tunnelling microscopes have established that charges in a doped Mott insulator have a tendency to cluster⁷, separating themselves from the antiferromagnetic background into which they have been introduced and sometimes forming charged 'stripes'. These stripes have been studied both experimentally⁸ and theoretically⁹, but the role they play in the superconducting state remains unclear.

The lack of a clear consensus on the mechanism behind high-temperature superconductivity is due, at least in part, to the difficulty of growing high-quality crystals with electron densities that vary smoothly across the range of interest. Chemically doped materials with different electron densities also have varying amounts of disorder, greatly complicating the interpretation of experiments that try to compare the properties of

systems with different doping levels. The role of disorder in the underdoped copper oxides is especially strong, perhaps because of the tendency for charge separation¹⁰.

The achievement of Schön *et al.*³ appears to remove these problems simultaneously, by making it possible to vary electron density continuously and reversibly, without introducing any external sources of disorder. They achieved this by building a simple¹¹ electronic device known as a field-effect transistor on top of a particularly simple layered copper oxide crystal. By establishing a voltage difference between a metallic electrode and the crystal, they can add or remove charge from the CuO₂ layers: positive voltage injects electrons, and a negative voltage injects holes. This allows them to vary both electron and hole charge densities over the full range of interest. The success of this technique relies on the high-quality copper oxide crystals they can grow using a technique known as molecular-beam epitaxy, and on the quality of the interfaces between the copper oxide material and the metallic electrode.

So, at the flick of a switch, Schön *et al.* can convert an insulating copper oxide compound into a superconductor. They find maximum superconducting-transition temperatures of 89 K for the hole-doped copper oxide and 34 K for the electron-doped one. These transition temperatures are not particularly high, but they demonstrate that the CuO₂ layers are similar to those generated by chemical doping. However, because the mobile electrons exist in a single two-dimensional layer, it will be difficult to use many of the experimental techniques used to study bulk samples of copper oxide superconductors, such as angle-resolved photoemission and neutron scattering. Conversely, the geometry of these samples may be better suited to other techniques. Tunnelling experiments should work easily, for example, and we should look for further inspiration from studies of two-dimensional semiconductors, where the ability to tune electron density using electrodes has been exploited for nearly 40 years. The challenge will be to design experiments that exploit the geometry of these tunable single-layer superconductors, opening the way to disorder-free observations probing the nature of superconductivity in the copper oxides. ■

Allan Hugh MacDonald is in the Department of Physics, University of Texas at Austin, Austin, Texas 78712, USA.

e-mail: macd@physics.utexas.edu

1. Bednorz, J. G. & Müller, K. A. *Z. Phys. B* **64**, 189–193 (1986).
2. Orenstein, J. & Millis, A. J. *Science* **288**, 468–474 (2000).
3. Schön, J. H. *et al. Nature* **414**, 434–436 (2001).
4. Anderson, P. W. *Science* **235**, 1196–1198 (1987).
5. Shen, Z. X. & Dessau, D. S. *Phys. Rep.* **253**, 2–162 (1995).
6. Ding, H. *et al. Nature* **382**, 51–54 (1996).
7. Emery, V. J. *et al. Phys. Rev. Lett.* **64**, 475–478 (1990).
8. Tranquada, J. *et al. Nature* **375**, 561–563 (1995).
9. Zaanen, J. *Science* **286**, 251–252 (1999).
10. Pan, S. H. *et al. Nature* **413**, 282–285 (2001).
11. Sigrist, T. *et al. Nature* **334**, 389–394 (1998).

Daedalus

Corrosive water

Marine creatures often make their shells or skeletons from calcium carbonate, yet the sea floor under deep water sees very little of it. In their long fall from the surface, the shells of these creatures lose carbonate to deep, high-pressure water, which dissolves much more carbon dioxide than surface water. This gives soluble calcium bicarbonate. Similarly, deep sea water dissolves much more oxygen than surface water, and high-pressure oxygen solution dissolves organic matter very efficiently.

All this implies that deep seawater is actually quite a corrosive medium, although rather slow in its action. This might explain the delayed failure of deep-sea cables, and the lack of organic detritus on the ocean floor as a whole. Over the aeons it has simply been dissolved in the water. The hydrogen sulphide of 'black smoker' springs must also have played its part.

It struck Daedalus that the mass of human rubbish now piled into landfills or inefficiently burnt could instead be dumped at sea. Even modern mariners have learnt that "over the side is over", so new ships strive to be self-contained. The rumour that the path of the great liners from Sydney to Tilbury could be traced by the crockery and cutlery they left on the ocean floor, to avoid washing it up, will never now be verified.

So DREADCO oceanographers are lowering assorted human rubbish in stout wire baskets to the ocean floor, to judge the rate of its decomposition. With luck, many food residues and organic nasties will be swiftly dissolved. So will some ceramics and plastics. The workers will closely study the way water penetrates into each object. The crucial components may be the metals themselves — do they dissolve or stay coherent? Many of the wooden ships sunk in numerous wars seem not to have left much metal behind them; cannon extracted from centuries-old sunken frigates are claimed to ignite when exposed at last to air. This suggests that metals undergo remarkable chemical changes when kept for centuries under water.

The DREADCO researchers hope to show that much of the world's rubbish — food, cans, bottles, paper and plastics — can be safely disposed of on the deep ocean floor. It would also be fitting that Greenpeace, which forced Shell to beach an outmoded oil platform instead of dumping it on the ocean floor, should be shown to be precisely wrong.

David Jones

Pulmonary nitric oxide in mountain dwellers

Populations living at high altitudes have an adaptive mechanism to offset hypoxia.

Nitric oxide is synthesized in the lungs to help regulate blood flow, and its levels have been found to drop in species native to low altitudes, including humans, upon acute exposure to reduced oxygen concentration^{1–3}. But we show here that exhalation of nitric oxide by chronically hypoxic populations of Tibetans living at 4,200 m and of Bolivian Aymara at 3,900 m is unexpectedly increased compared with a low-altitude reference sample from the United States. This consistent response in two far-removed, high-altitude locales indicates that increasing the concentration of nitric oxide in the lungs may represent a means of offsetting hypoxia.

We measured the geometric mean concentration of nitric oxide (NO) exhaled by individuals in a group of healthy non-smokers from Tibet (Fig. 1) as 18.6 p.p.b. ($n = 105$; range, 5.5–55.7 p.p.b.; coefficient of variation (c.v.), 1.5%) — more than twice the value for the low-altitude reference sample (7.4 p.p.b.; $n = 33$; range, 4.5–14.6; c.v., 2.4%). The geometric mean NO concentration exhaled by the Aymara was 9.5 p.p.b. ($n = 144$; range, 2.7–30.3; c.v., 1.9%) — over 25% greater than that for the low-altitude reference (Fig. 2). Artificial relief from hypoxia (inspiration of oxygen at 42–50% v/v concentration) resulted in an increase of 2.5 p.p.b. in NO exhaled by Tibetans ($n = 26$, $P < 0.05$), but caused no change among the Aymara ($n = 25$, $P > 0.05$), suggesting that the mechanism for sustaining high NO levels may differ between the two populations.

Measurements of NO exhaled at the mouth reflect accurately and qualitatively the dynamics of NO production and consumption in the lungs¹. Reduced consumption is unlikely to explain the higher NO values of high-altitude natives, because haemoglobin rapidly scavenges NO and both samples showed high haemoglobin concentrations compared with those taken from sea-level dwellers (see supplementary information).



Figure 1 The high altitude at which Tibetans live means that they are constantly exposed to reduced atmospheric oxygen concentrations.

Reduced consumption is also unlikely to explain the higher NO levels found in Tibetans compared with the Aymara.

Controlling for the significantly higher haemoglobin concentrations in Aymara samples by comparing subsamples with the same range of haemoglobin (153–158 g l⁻¹) revealed mean NO concentrations of 19.9 p.p.b. and 8.8 p.p.b. for the 16 Tibetans and the 25 Aymara, respectively. NO levels did not correlate with haemoglobin concentration or resting oxygen saturation in either sample. An increase in NO synthesis is therefore a more likely explanation for the high NO concentrations in these two samples (such an increase would therefore be smaller in the Aymara).

Nitric oxide is synthesized by NO-synthase enzymes and its synthesis depends on the availability of molecular oxygen^{1,4,5}, so a drop in oxygen concentration would be expected to result in reduced NO production by synthases. Possible adaptations to maintain high-output NO synthesis under hypoxia include variant forms of the

enzyme that show altered kinetics for oxygen dependence, modification of co-factor availability through post-translational effects, and/or increased expression of the synthase enzymes themselves.

To assess the potential benefits of increased levels of endogenous NO to high-altitude populations, we investigated whether increasing NO at sea level improves oxygen uptake by the lungs and thereby offsets hypoxia. Oxygen uptake improved in a dose-dependent manner ($P < 0.05$) in the presence of exogenous NO at concentrations of 1.3–31,600 p.p.b. during hypoxic ventilation (10.4% oxygen; measured by the progressive reduction in oxygen concentration at the end of a 10-s exhalation, when values reflect exhalate concentration from the alveoli where gas exchange occurs). End-exhalation oxygen concentration decreased from 8.5% to 7.8% as exogenous NO concentration was increased; over 70% of the effect on oxygen uptake occurred in the physiological range at 290 p.p.b. NO. An unchanged end-exhalation CO₂ measurement confirmed that increased oxygen uptake was not due to a momentary increase in oxygen consumption or to metabolic demands.

The human model of high-altitude adaptation should also incorporate a functional, adaptive benefit of high NO levels in the lungs. NO produced in the lungs dilates pulmonary blood vessels, increases pulmonary blood flow and reduces pulmonary hypertension. By reacting with haemoglobin in red blood cells, NO increases haemoglobin oxygenation and may improve the delivery of oxygen to tissues by enhancing systemic vasodilation and blood flow⁶.

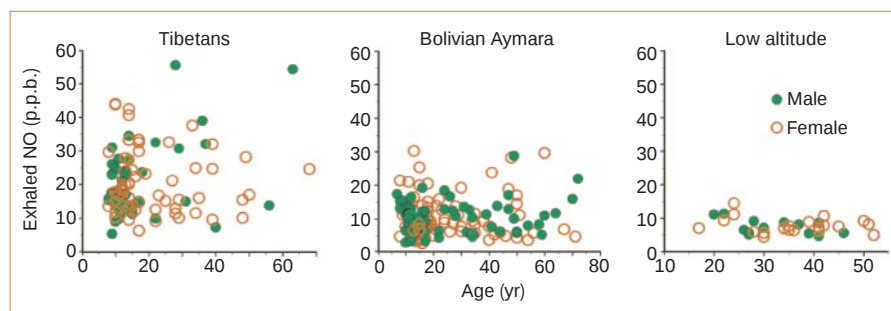


Figure 2 A Tibetan population living at 4,200 m, a Bolivian Aymara population at 3,900 m and a low-altitude population in the United States differ significantly in their mean concentrations of exhaled nitric oxide (NO; ANOVA, $F = 77.9$, d.f. = 2, $P < 0.05$); no sex or age differences are evident in the results. Details of methods are available from the authors.

Several distinctive features of Tibetan and Andean high-altitude natives are related to their increased vasodilation and blood flow compared with acclimatized lowlanders. For example, Tibetans are better than Han Chinese at increasing their cerebral blood flow during exercise⁷, as well as their utero-placental blood flow⁸; the Andean Aymara show a large capacity for pulmonary diffusion⁹ and the Andean Quechua have better circulation to cold extremities than Europeans¹⁰; and Tibetan and Andean natives show higher oxygen saturation during exercise than do acclimatized Han Chinese and Europeans^{11,12}.

The similar responses of these two geographically separate high-altitude populations underlines the importance of NO for life under hypoxic stress. The functional advantage of high NO concentrations in the lungs seems to be to offset ambient hypoxia by enhancing the uptake of oxygen from the lungs, which presumably improves delivery of oxygen to peripheral tissues.

Cynthia M Beall*, **Daniel Laskowski†**, **Kingman P. Strohl‡**, **Rudy Soria§**, **Mercedes Villena§**, **Enrique Vargas§**, **Ana Maria Alarcon§**, **Cristina Gonzales§**, **Serpil C. Erzurum†**

*Department of Anthropology, Case Western Reserve University, Cleveland, Ohio 44106, USA

e-mail: cmb2@po.cwru.edu

†Departments of Pulmonary and Critical Care Medicine, and Cancer Biology, Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA

‡Department of Medicine, Louis Stokes VA Medical Center, Case Western Reserve University, Cleveland, Ohio 44106, USA

§Departamento Respiratorio, Instituto Boliviano de Biología de Altura, La Paz, Bolivia

1. Dweik, R. A. et al. *J. Clin. Invest.* **101**, 660–666 (1998).
2. Le Cras, T. D. & McMurtry, I. F. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **280**, L575–L582 (2001).
3. Gustafsson, L. E., Leone, A. M., Persson, M. G., Wiklund, N. P. & Moncada, S. *Biochem. Biophys. Res. Commun.* **181**, 852–857 (1991).
4. Abu-Soud, H., Rousseau, D. & Stuehr, D. J. *Biol. Chem.* **271**, 32515–32518 (1996).
5. Hurshman, A. & Marletta, M. *Biochemistry* **34**, 5627–5634 (1995).
6. Gow, A. J., Luchsinger, B. P., Pawloski, J. R., Singel, D. J. & Stamler, J. S. *Proc. Natl Acad. Sci. USA* **96**, 9027–9032 (1999).
7. Huang, S. Y. et al. *J. Appl. Physiol.* **73**, 2638–2642 (1992).
8. Moore, L. G., Zamudio, S., Zhuang, J., Sun, S. & Droma, T. *Am. J. Phys. Anthropol.* **114**, 42–53 (2001).
9. Spielvogel, H., Vargas, E., Antezana, G., Barragan, L. & Cudkovic, L. *Respiration* **35**, 125–135 (1978).
10. Little, M. in *Man in the Andes: A Multidisciplinary Study of High-Altitude Quechua* (eds Baker, P. & Little, M.) 332–362 (Downen, Hutchinson & Ross, Stroudsburg, Pennsylvania, 1976).
11. Sun, S. F. et al. *Resp. Physiol.* **79**, 151–162 (1990).
12. Brutsaert, T. D., Araoz, M., Soria, R., Spielvogel, H. & Haas, J. E. *Am. J. Phys. Anthropol.* **113**, 169–182 (2000).

Supplementary information accompanies this communication on Nature's website (www.nature.com).

Competing financial interests: declared none.

Longevity

Extending the lifespan of long-lived mice

Ames dwarf mice are mutant mice that live about 50% longer than their normal siblings^{1–3} because they carry a 'longevity' gene, *Prop1*^{df}, and in some phenotypic respects they resemble normal mice whose lifespan has been extended by restricted food intake^{2,4,5}. Here we investigate whether these factors influence lifespan by similar or independent mechanisms, by deliberately reducing the number of calories consumed by Ames dwarf mice. We show that calorie restriction confers a further lifespan increase in the dwarfs, indicating that the two factors may act through different pathways.

To investigate the effects of calorie restriction on the already extended lifespan of Ames dwarf mice, we divided 45 2-month-old Ames dwarf mice and 53 of their normal siblings into two groups, which were subjected either to calorie restriction (CR) or to continued feeding *ad libitum* (AL). We fed CR mice daily, reducing their food intake in successive weeks to 90%, 80% and finally 70% of that consumed daily by genotype- and sex-matched AL animals⁶. Because the food consumption of AL mice declines naturally with

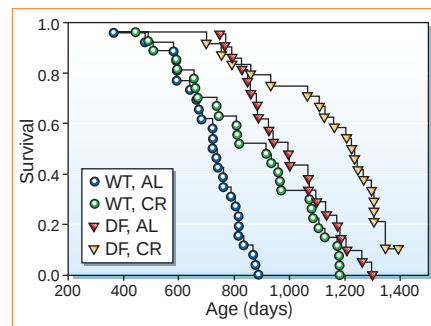


Figure 1 Survival plots of Ames dwarf (DF) and normal (wild-type, WT) mice fed *ad libitum* (AL) or restricted to 70% of normal calorie intake (calorie restriction, CR).

age, the amount of food given to CR animals was kept constant after the age of 2 years.

The survival curves shown in Fig. 1 indicate that calorie restriction causes a further significant increase in the longevity of Ames dwarf mice. When males and females are considered together, the difference between the CR and AL groups of Ames dwarf mice is significant ($P < 0.004$, log rank test). The effect of calorie restriction on lifespan in Ames dwarf mice is also significant ($P < 0.05$) when genders are considered separately. As expected, calorie restriction also extends the lifespan of normal mice ($P < 0.002$), although AL Ames dwarf mice outlive AL normal mice ($P < 0.00001$). Moreover, CR Ames dwarf

mice outlive CR normal mice ($P < 0.0001$).

The survival plots (Fig. 1) reveal a further disparity: although both dwarfism and calorie restriction extend longevity, the effect of reduced food intake is associated primarily with a change in the slope of the survival curve (that is, it reduces the rate of age-related mortality), whereas the effect of dwarfism mainly reflects a shift in the age at which the age-dependent increase in mortality risk first becomes appreciable. Calorie restriction therefore seems to decelerate ageing, whereas the *Prop1*^{df} allele seems to delay it.

Our results indicate that long-lived Ames dwarf mice are not merely mimics of CR mice, and show that the pathways responsible for extending lifespan in the dwarfs and in CR animals are not identical. However, features that are shared by CR normal mice and Ames dwarf mice, and by long-lived knockout mice that lack the growth-hormone receptor⁷, include reduced body size and lower plasma levels of insulin, the insulin-like growth factor IGF-1, glucose and thyroid hormone. These factors may contribute to delayed ageing and increased longevity in each of these animal models.

For example, the IGF/insulin or a similar signalling pathway is involved in lifespan determination in the fruitfly *Drosophila melanogaster*^{8,9}, the roundworm *Caenorhabditis elegans*¹⁰, and yeast¹¹. This supports the idea that hormonal regulation of metabolic pathways in response to altered food availability may be a way of regulating lifespan that is deeply rooted in evolutionary history.

Andrzej Bartke*, **J. Chris Wright***, **Julie A. Mattison†**, **Donald K. Ingram†**, **Richard A. Miller‡**, **George S. Roth§**

*Department of Physiology, School of Medicine, Southern Illinois University, Carbondale, Illinois 62901, USA

e-mail: abartke@siumed.edu

†National Institute on Aging, NIH Animal Center, 16701 Elmer School Road, Poolesville, Maryland 20837, USA

‡University of Michigan, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109, USA

§Gerontology Research Center, 5600 Nathan Shock Drive, Baltimore, Maryland 21224, USA

1. Brown-Borg, H. M., Borg, K. E., Meliska, C. J. & Bartke, A. *Nature* **384**, 33 (1996).
2. Bartke, A. in *The Molecular Genetics of Aging* (ed. Hekimi, S.) Vol. 29, 181–202 (Springer, New York, 2000).
3. Flurkey, K., Papaconstantinou, J., Miller, R. A. & Harrison, D. E. *Proc. Natl Acad. Sci. USA* **98**, 6736–6741 (2001).
4. Harris, S. B., Gunion, M. W., Rosenthal, M. J. & Walford, R. L. *Mech. Ageing Dev.* **73**, 209–221 (1994).
5. Weindruch, R. *Toxicol. Pathol.* **24**, 742–745 (1996).
6. Mattison, J. A. et al. *J. Am. Aging Assoc.* **23**, 9–16 (2000).
7. Coschigano, K. T., Clemmons, D., Bellush, L. L. & Kopchick, J. J. *Endocrinology* **141**, 2608–2613 (2000).
8. Clancy, D. J. et al. *Science* **292**, 104–106 (2001).
9. Tatar, M. et al. *Science* **292**, 107–110 (2001).
10. Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *Science* **277**, 942–946 (1997).
11. Fabrizio, P., Pozza, F., Pletcher, S. D., Gendron, C. M. & Longo, V. D. *Science* **292**, 288–290 (2001).

Long-distance quantum communication with atomic ensembles and linear optics

L.-M. Duan^{*†}, M. D. Lukin[‡], J. I. Cirac^{*} & P. Zoller^{*}

^{*} Institut für Theoretische Physik, Universität Innsbruck, A-6020 Innsbruck, Austria

[†] Laboratory of Quantum Communication and Computation, USTC, Hefei 230026, China

[‡] Physics Department and ITAMP, Harvard University, Cambridge, Massachusetts 02138, USA

Quantum communication holds promise for absolutely secure transmission of secret messages and the faithful transfer of unknown quantum states. Photonic channels appear to be very attractive for the physical implementation of quantum communication. However, owing to losses and decoherence in the channel, the communication fidelity decreases exponentially with the channel length. Here we describe a scheme that allows the implementation of robust quantum communication over long lossy channels. The scheme involves laser manipulation of atomic ensembles, beam splitters, and single-photon detectors with moderate efficiencies, and is therefore compatible with current experimental technology. We show that the communication efficiency scales polynomially with the channel length, and hence the scheme should be operable over very long distances.

The goal of quantum communication is to transmit quantum states between distant sites. Such transmission has potential application in the secret transfer of classical messages by means of quantum cryptography¹, and is also an essential element in the construction of quantum networks. The basic problem of quantum communication is to generate nearly perfect entangled states between distant sites. Such states can be used, for example, to implement secure quantum cryptography using the Ekert protocol¹, and to faithfully transfer quantum states via quantum teleportation². All realistic schemes for quantum communication are at present based on the use of photonic channels. However, the degree of entanglement generated between two distant sites normally decreases exponentially with the length of the connecting channel, because of optical absorption and other channel noise. To regain a high degree of entanglement, purification schemes can be used³, but this does not fully solve the long-distance communication problem. Because of the exponential decay of the entanglement in the channel, an exponentially large number of partially entangled states are needed to obtain one highly entangled state, which means that for a sufficiently long distance the task becomes nearly impossible.

To overcome the difficulty associated with the exponential fidelity decay, the concept of quantum repeaters can be used⁴. In principle, this allows the overall communication fidelity to be made very close to unity, with the communication time growing only polynomially with transmission distance. In analogy to fault-tolerant quantum computing^{5,6}, the proposed quantum repeater is a cascaded entanglement-purification protocol for communication systems. The basic idea is to divide the transmission channel into many segments, with the length of each segment comparable to the channel attenuation length. First, entanglement is generated and purified for each segment; the purified entanglement is then extended to a greater length by connecting two adjacent segments through entanglement swapping^{2,7}. After this swapping, the overall entanglement is decreased, and has to be purified again. The rounds of entanglement swapping and purification can be continued until nearly perfect entangled states are created between two distant sites.

To implement the quantum repeater protocol, we need to generate entanglement between distant quantum bits (qubits), store them for a sufficiently long time and perform local collective operations on several of these qubits. Quantum memory is essential, because all purification protocols are probabilistic. When entanglement purification is performed for each segment of the channel, quantum memory can be used to keep the segment state if the purification succeeds, and to repeat the purification for the

segments only where the previous attempt fails. This is essential for ensuring polynomial scaling in the communication efficiency, because if there were no available memory, the purifications for all the segments would need to succeed at the same time; the probability of such an event decreases exponentially with channel length. The requirement of quantum memory implies that we need to store the local qubits in atomic internal states instead of photonic states, as it is difficult to store photons for a reasonably long time. With atoms as the local information carriers, it seems to be very hard to implement quantum repeaters: normally, one needs to achieve the strong coupling between atoms and photons by using high-finesse cavities for atomic entanglement generation, purification, and swapping^{8,9}, which, in spite of recent experimental advances^{10–12}, remains a very challenging technology.

Here we propose a different scheme, which realizes quantum repeaters and long-distance quantum communication with simple physical set-ups. The scheme is a combination of three significant advances in entanglement generation, connection, and applications, with each of the steps having built-in entanglement purification and resilience to realistic noise. The scheme for fault-tolerant entanglement generation originates from earlier proposals to entangle single atoms through single-photon interference at photodetectors^{13,14}. But the present approach involves collective excitations in atomic ensembles rather than in single particles, which allows simpler realization and greatly improved generation efficiency. This is due to collectively enhanced coupling to light, which has been recently investigated both theoretically^{15–19} and experimentally^{20–22}. The entanglement connection is achieved through simple linear optical operations, and is inherently robust against realistic imperfections. Different schemes with linear optics have been proposed recently for quantum computation²³ and purification²⁴. Finally, the resulting state of ensembles after the entanglement connection finds direct applications in realizing entanglement-based quantum communication protocols, such as quantum teleportation, cryptography, and Bell inequality detection. In all of these applications, the mixed entanglement is purified automatically to nearly perfect entanglement. As a combination of these three advances, our scheme circumvents the realistic noise and imperfections, and provides a feasible method of long-distance high-fidelity quantum communication. The required overhead in communication time increases with distance only polynomially.

Entanglement generation

The basic element of our system is a cloud of N_a identical atoms with

the relevant level structure shown in Fig. 1. A pair of metastable lower states $|g\rangle$ and $|s\rangle$ can correspond to—for example—hyperfine or Zeeman sublevels of the electronic ground state of alkali-metal atoms. Long lifetimes for the relevant coherence have been observed both in a room-temperature dilute atomic gas (see, for example, ref. 21) and in a sample of cold trapped atoms (see, for example, refs 20, 22). To facilitate enhanced coupling to light, the atomic medium is preferably optically thick along one direction. This can be achieved either by working with a pencil-shaped atomic sample^{20–22} or by placing the sample in a low-finesse ring cavity^{17,25} (see Supplementary Information).

All the atoms are initially prepared in the ground state $|g\rangle$. A sample is illuminated by a short, off-resonant laser pulse that induces Raman transitions into the states $|s\rangle$. We are particularly interested in the forward-scattered Stokes light that is co-propagating with the laser. Such scattering events are uniquely correlated with the excitation of the symmetric collective atomic mode S (refs 15–22) given by $S \equiv (1/\sqrt{N_a})\sum_i |g_i\rangle\langle s_i|$, where the summation is taken over all the atoms. In particular, an emission of the single Stokes photon in a forward direction results in the state of atomic

ensemble given by $S^\dagger|0_a\rangle$, where the ensemble ground state $|0_a\rangle \equiv \bigotimes_i |g_i\rangle$.

We assume that the light–atom interaction time t_Δ is short, so that the mean photon number in the forward-scattered Stokes pulse is much smaller than 1. We can define an effective single-mode bosonic operator a for this Stokes pulse with the corresponding vacuum state denoted by $|0_p\rangle$. The whole state of the atomic collective mode and the forward-scattering Stokes mode can now be written in the following form (see Supplementary Information for details)

$$|\phi\rangle = |0_a\rangle|0_p\rangle + \sqrt{p_c}S^\dagger a^\dagger|0_a\rangle|0_p\rangle + o(p_c) \quad (1)$$

where p_c is the small excitation probability, and $o(p_c)$ represents the terms with more excitations whose probabilities are equal to or smaller than p_c^2 . Before proceeding, we note that a fraction of light is emitted in other directions owing to spontaneous emissions. But whenever N_a is large, the contribution from the spontaneous emissions to the population in the symmetric collective mode is small^{15–22}. As a result, we have a large signal-to-noise ratio for the processes involving the collective mode, which greatly enhances the efficiency of the present scheme (see Box 1 and Supplementary Information).

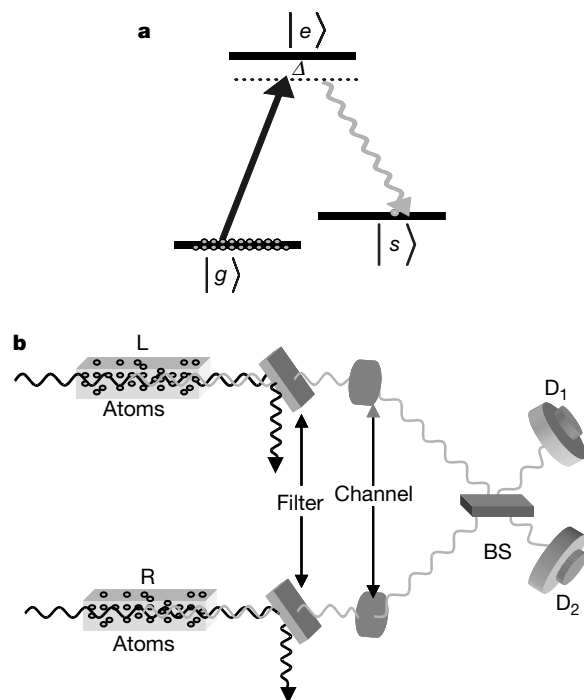


Figure 1 Set-up for entanglement generation. **a**, The relevant level structure of the atoms in the ensemble, with $|g\rangle$, the ground state, $|s\rangle$, the metastable state for storing a qubit, and $|e\rangle$, the excited state. The transition $|g\rangle \rightarrow |e\rangle$ is coupled by the classical laser (the pumping light) with the Rabi frequency Ω , and the forward-scattered Stokes light comes from the transition $|e\rangle \rightarrow |s\rangle$, which has a different polarization and frequency to the pumping light. For convenience, we assume off-resonant coupling with a large detuning Δ . **b**, Schematic set-up for generating entanglement between the two atomic ensembles L and R. The two ensembles are pencil-shaped, and illuminated by the synchronized classical pumping pulses. The forward-scattered Stokes pulses are collected and coupled to optical channels (such as fibres) after the filters, which are polarization- and frequency-selective to filter the pumping light. The pulses after the transmission channels interfere at a 50%-50% beam splitter BS, with the outputs detected respectively by two single-photon detectors D_1 and D_2 . If there is a click in D_1 or D_2 , the process is finished and we successfully generate entanglement between the ensembles L and R. Otherwise, we first apply a repumping pulse (to the transition $|s\rangle \rightarrow |e\rangle$) to the ensembles L and R, to set the state of the ensembles back to the ground state $|0_a\rangle \equiv \bigotimes_i |g_i\rangle$, then the same classical laser pulses as the first round are applied to the transition $|g\rangle \rightarrow |e\rangle$ and we detect again the forward-scattered Stokes pulses after the beam splitter. This process is repeated until finally we have a click in the D_1 or the D_2 detector.

Box 1

Collective enhancement

Long-lived excitations in atomic ensembles can be viewed as waves of excited spins. We are here particularly interested in the symmetric spin wave mode S . For a simple demonstration of collective enhancement, we assume that the atoms are placed in a low-finesse ring cavity²⁵, with a relevant cavity mode corresponding to forward-scattered Stokes radiation. The cavity-free case corresponds to the limit where the finesse tends to 1 (ref. 17). The interaction between the forward-scattered light mode and the atoms is described by the hamiltonian

$$H = \hbar \left(\sqrt{N_a} \Omega g_c \Delta \right) S^\dagger b^\dagger + \text{h.c.}$$

where h.c. is the hermitian conjugation, $b^\dagger$ is the creation operator for cavity photons, Ω is the laser Rabi frequency, and g_c the atom–field coupling constant. In addition to coherent evolution, the photonic field mode can leak out of the cavity at a rate κ , whereas atomic coherence is dephased by spontaneous photon scattering into random directions that occurs at a rate $\gamma'_s = \Omega^2/\Delta^2 \gamma_s$ for each atom, with γ_s being the natural linewidth of the electronic excited state. We emphasize that in the absence of superradiant effects, spontaneous emission events are independent for each atom.

In the bad-cavity limit, we can adiabatically eliminate the cavity mode, and the resulting dynamics for the collective atomic mode is described by the Heisenberg–Langevin equation (see Supplementary Information for details)

$$\dot{S}^\dagger = \frac{(\kappa' - \gamma'_s)}{2} S^\dagger - \sqrt{\kappa'} b_{\text{in}}(t) + \text{noise}$$

where $\kappa' = 4|\Omega|^2 g_c^2 N_a / (\Delta^2 \kappa)$, b_{in} is a vacuum field leading into the cavity, and the last term represents the fluctuating noise field corresponding to spontaneous emission. We note that the nature of the dynamics is determined by the ratio between the build-up of coherence due to forward-scattered photons κ' and coherence decay due to spontaneous emission γ'_s . The signal-to-noise ratio is therefore given by $R = \kappa'/\gamma'_s = 4N_a g_c^2 / (\kappa \gamma_s)$, which is large when a many-atom ensemble is used. In the cavity-free case, this expression corresponds to the optical depth (density-length product) of the sample. The result should be compared with the signal-to-noise ratio in the single-atom case $N_a = 1$, where to obtain $R > 1$ a high- Q microcavity is required^{10–12}. The collective enhancement takes place because the coherent forward scattering involves only one collective atomic mode S , whereas the spontaneous emissions distribute excitation over all atomic modes. Therefore only a small fraction of spontaneous emission events influences the symmetric mode S , which results in a large signal-to-noise ratio.

We now show how to use this set-up to generate entanglement between two distant ensembles L (left) and R (right) using the configuration shown in Fig. 1. Here two laser pulses excite both ensembles simultaneously, and the whole system is described by the state $|\phi\rangle_L \otimes |\phi\rangle_R$, where $|\phi\rangle_L$ and $|\phi\rangle_R$ are given by equation (1) with all the operators and states distinguished by the subscript L or R. The forward-scattered Stokes light from both ensembles is combined at the beam splitter, and a photodetector click in either D_1 or D_2 measures the combined radiation from two samples, $a_+^\dagger a_+$ or $a_-^\dagger a_-$ with $a_\pm = (a_L \pm e^{i\varphi} a_R)/\sqrt{2}$. Here, φ denotes an unknown difference of the phase shifts in the left and the right side channels. We can also assume that φ has an imaginary part to account for the possible asymmetry of the set-up, which will also be corrected automatically in our scheme. But the set-up asymmetry can be easily made very small, and for simplicity of expressions we assume that φ is real in the following. Conditional on the detector click, we should apply a_+ or a_- to the whole state $|\phi\rangle_L \otimes |\phi\rangle_R$, and the projected state of the ensembles L and R is nearly maximally entangled, with the form (neglecting the high-order terms $o(p_c)$):

$$|\Psi_{\varphi/LR}^\pm\rangle = (S_L^\dagger \pm e^{i\varphi} S_R^\dagger)/\sqrt{2} |0_a\rangle_L |0_a\rangle_R \quad (2)$$

The probability of getting a click is given by p_c for each round, so we need to repeat the process about $1/p_c$ times for a successful entanglement preparation, and the average preparation time is given by $T_0 \approx t_d/p_c$. The states $|\Psi_{\varphi/LR}^+\rangle$ and $|\Psi_{\varphi/LR}^-\rangle$ can be transformed to each other by a simple local phase shift. Without loss of generality, we assume in the following that we generate the entangled state $|\Psi_{\varphi/LR}^+\rangle$.

As will be shown below, the presence of noise modifies the projected state of the ensembles to

$$\rho_{LR}(c_0, \varphi) = \frac{1}{c_0 + 1} (c_0 |0_a 0_a\rangle_{LR} \langle 0_a 0_a| + |\Psi_{\varphi/LR}^+\rangle \langle \Psi_{\varphi/LR}^+|) \quad (3)$$

where the ‘vacuum’ coefficient c_0 is determined by the dark count rates of the photon detectors. It will be seen below that any state in the form of equation (3) will be purified automatically to a maximally entangled state in the entanglement-based communication schemes. We therefore call this state an effective maximally entangled (EME) state, with the vacuum coefficient c_0 determining the purification efficiency.

Entanglement connection through swapping

After the successful generation of entanglement within the

attenuation length, we want to extend the quantum communication distance. This is done through entanglement swapping with the configuration shown in Fig. 2. Suppose that we start with two pairs of entangled ensembles described by the state $\rho_{L_1} \otimes \rho_{L_2}$, where ρ_{L_1} and ρ_{L_2} are given by equation (3). In the ideal case, the set-up shown in Fig. 2 measures the quantities corresponding to operators $S_\pm^\dagger S_\pm$ with $S_\pm = (S_{11} \pm S_{22})/\sqrt{2}$. If the measurement is successful (that is, one of the detectors registers one photon), we will prepare the ensembles L and R into another EME state. The new φ -parameter is given by $\varphi_1 + \varphi_2$, where φ_1 and φ_2 denote the old φ -parameters for the two segment EME states. As will be seen below, even in the presence of realistic noise and imperfections, an EME state is still created after a detector click. The noise only influences the success probability of getting a click and the new vacuum coefficient in the EME state. In general, we can express the success probability p_i and the new vacuum coefficient c_i as $p_i = f_1(c_0)$ and $c_i = f_2(c_0)$, where the functions f_1 and f_2 depend on the particular noise properties.

The above method for connecting entanglement can be cascaded to arbitrarily extend the communication distance. For the i th ($i = 1, 2, \dots, n$) entanglement connection, we first prepare in parallel two pairs of ensembles in the EME states with the same vacuum coefficient c_{i-1} and the same communication length L_{i-1} , and then perform entanglement swapping as shown in Fig. 2, which now succeeds with a probability $p_i = f_1(c_{i-1})$. After a successful detector click, the communication length is extended to $L_i = 2L_{i-1}$, and the vacuum coefficient in the connected EME state becomes $c_i = f_2(c_{i-1})$. As the i th entanglement connection needs to be repeated on average $1/p_i$ times, the total time needed to establish an EME state over the distance $L_n = 2^n L_0$ is given by $T_n = T_0 \prod_{i=1}^n (1/p_i)$, where L_0 denotes the distance of each segment in the entanglement generation.

Entanglement-based communication schemes

After an EME state has been established between two distant sites, we would like to use it in communication protocols, such as quantum teleportation, cryptography, and Bell inequality detection. It is not obvious that the EME state of equation (3), which is entangled in the Fock basis, is useful for these tasks, as in the Fock basis it is experimentally hard to do certain single-bit operations. We will now show how the EME states can be used to realize all these protocols with simple experimental configurations.

Quantum cryptography and Bell inequality detection are achieved with the set-up shown by Fig. 3a. The state of the two

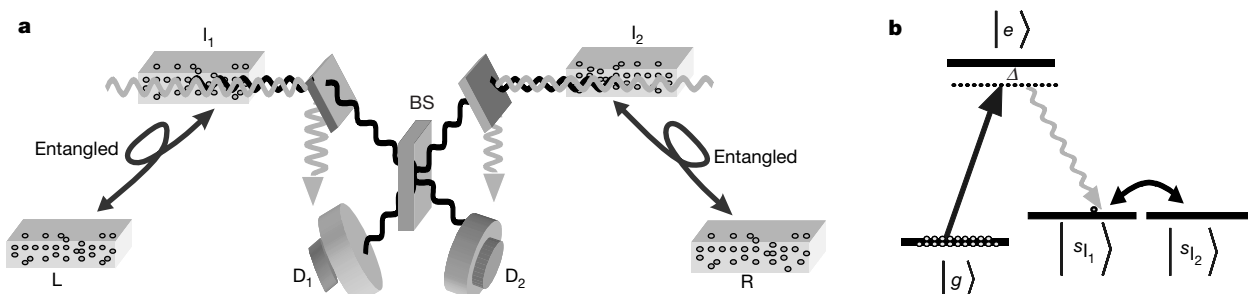


Figure 2 Set-up for entanglement connection. **a**, Illustrative set-up for the entanglement swapping. We have two pairs of ensembles—L and I_1 , and I_2 and R—distributed at three sites L, I and R. Each of the ensemble-pairs L and I_1 , and I_2 and R is prepared in an EME state in the form of equation (3). The stored atomic excitations of two nearby ensembles I_1 and I_2 are converted simultaneously into light. This is achieved by applying a retrieval pulse of suitable polarization that is near-resonant with the atomic transition $|s\rangle \rightarrow |e\rangle$, which causes coherent conversion of atomic excitations into photons that have a different polarization and frequency to the retrieval pulse^{18,21,22}. The efficiency of this transfer can be very close to unity even at a single quantum level owing to collective enhancement^{18,21,22}. After the transfer, the stimulated optical excitations interfere at a

50%-50% beam splitter, and then detected by the single-photon detectors D_1 and D_2 . If either D_1 or D_2 clicks, the protocol is successful and an EME state in the form of equation (3) is established between the ensembles L and R with a doubled communication distance. Otherwise, the process fails, and we need to repeat the previous entanglement generation and swapping until finally we have a click in D_1 or D_2 , that is, until the protocol finally succeeds. **b**, The two intermediate ensembles I_1 and I_2 can also be replaced by one ensemble but with two metastable states I_1 and I_2 to store the two different collective modes. The 50%-50% beam splitter operation can be simply realized by a $\pi/2$ pulse applied to the two metastable states before the collective atomic excitations are transferred to the optical excitations.

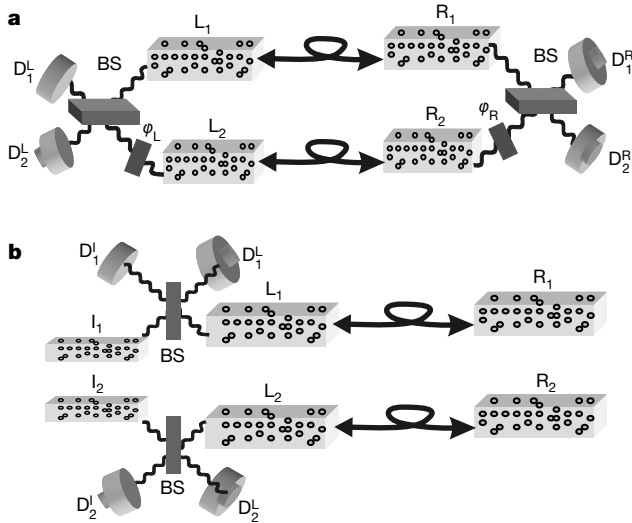


Figure 3 Set-up for entanglement-based communication schemes. **a**, Schematic set-up for the realization of quantum cryptography and Bell inequality detection. Two pairs of ensembles L_1, R_1 and L_2, R_2 (or two pairs of metastable states as shown in **b**) have been prepared in the EME states. The collective atomic excitations on each side are transferred to the optical excitations, which, respectively after a relative phase shift φ_L or φ_R and a 50%-50% beam splitter, are detected by the single-photon detectors D_1^L, D_2^L and D_1^R, D_2^R . We look at the four possible coincidences of D_1^R, D_2^R with D_1^L, D_2^L , which are functions of the phase difference $\varphi_L - \varphi_R$. Depending on the choice of φ_L and φ_R , this set-up can realize both the quantum cryptography and the Bell inequality detection. **b**, Schematic set-up for probabilistic quantum teleportation of the atomic 'polarization' state. Similarly, two pairs of ensembles L_1, R_1 and L_2, R_2 are prepared in the EME states. We want to teleport an atomic 'polarization' state $(d_0 S_{L_1}^\dagger + d_1 S_{L_2}^\dagger)|0_a 0_{aL_1 L_2}\rangle$ with unknown coefficients d_0, d_1 from the left to the right side, where $S_{L_1}^\dagger, S_{L_2}^\dagger$ denote the collective atomic operators for the two ensembles L_1 and L_2 (or two metastable states in the same ensemble). The collective atomic excitations in the ensembles L_1, L_2 and L_1, L_2 are transferred to the optical excitations, which, after a 50%-50% beam splitter, are detected by the single-photon detectors D_1^L, D_2^L and D_1^R, D_2^R . If, and only if, there is one click in D_1^L, D_2^L and one click in D_1^R, D_2^R , the protocol is successful. When the protocol succeeds, the collective excitation in the ensembles R_1 and R_2 , if appearing, would be found in the same 'polarization' state $(d_0 S_{R_1}^\dagger + d_1 S_{R_2}^\dagger)|0_a 0_{aR_1 R_2}\rangle$ up to a local π -phase rotation.

pairs of ensembles is expressed as $\rho_{L_1 R_1} \otimes \rho_{L_2 R_2}$, where $\rho_{L_i R_i}$ ($i = 1, 2$) denote the same EME state with the vacuum coefficient c_n if we have done entanglement connection n times. The φ -parameters in $\rho_{L_1 R_1}$ and $\rho_{L_2 R_2}$ are the same, provided that the two states are established over the same stationary channels. We register only the coincidences of the two-side detectors, so the protocol is successful only if there is a click on each side. Under this condition, the vacuum components in the EME states, together with the state components $S_{L_1}^\dagger S_{L_2}^\dagger |\text{vac}\rangle$ and $S_{R_1}^\dagger S_{R_2}^\dagger |\text{vac}\rangle$, where $|\text{vac}\rangle$ denotes the ensemble state $|0_a 0_a 0_a\rangle_{L_1 R_1 L_2 R_2}$, have no contributions to the experimental results. So, for the measurement scheme shown by Fig. 3, the ensemble state $\rho_{L_1 R_1} \otimes \rho_{L_2 R_2}$ is effectively equivalent to the following 'polarization' maximally entangled (PME) state (the terminology of 'polarization' comes from an analogy to the optical case):

$$|\Psi\rangle_{\text{PME}} = (S_{L_1}^\dagger S_{R_2}^\dagger + S_{L_2}^\dagger S_{R_1}^\dagger)/\sqrt{2} |\text{vac}\rangle \quad (4)$$

The success probability for the projection from $\rho_{L_1 R_1} \otimes \rho_{L_2 R_2}$ to $|\Psi\rangle_{\text{PME}}$ (that is, the probability of getting a click on each side) is given by $p_a = 1/[2(c_n + 1)^2]$. We can also check that in Fig. 3, the phase shift φ_A ($A = L$ or R) together with the corresponding beam-splitter operation are equivalent to a single-bit rotation in the basis $\{|0\rangle_A \equiv S_{A_1}^\dagger |0_a 0_a\rangle_{A_1 A_2}, |1\rangle_A \equiv S_{A_2}^\dagger |0_a 0_a\rangle_{A_1 A_2}\}$ with the rotation angle $\theta = \varphi_A/2$. Now it is clear how to do quantum cryptography and Bell inequality detection, as we have the PME state and we can

perform the desired single-bit rotations in the corresponding basis. For instance, to distribute a quantum key between the two remote sides, we simply choose φ_A randomly from the set $\{0, \pi/2\}$ with an equal probability, and keep the measurement results (to be 0 if D_1^A clicks, and 1 if D_2^A clicks) on both sides as the shared secret key if the two sides become aware that they have chosen the same phase shift after the public declaration of φ_A . This is exactly the Ekert scheme¹, and its absolute security follows directly from the proofs in refs 26 and 27. For the Bell inequality detection, we infer the correlations $E(\varphi_L, \varphi_R) \equiv P_{D_1^L D_1^R} + P_{D_2^L D_2^R} - P_{D_1^L D_2^R} - P_{D_2^L D_1^R} = \cos(\varphi_L - \varphi_R)$ from the measurement of the coincidences $P_{D_1^L D_1^R}$ and so on. For the set-up shown in Fig. 3a, we would have $|E(0, \pi/4) + E(\pi/2, \pi/4) + E(\pi/2, 3\pi/4) - E(0, 3\pi/4)| = 2\sqrt{2}$, whereas for any local hidden variable theories, the CHSH inequality²⁸ implies that this value should be below 2.

We can also use the established long-distance EME states for faithful transfer of unknown quantum states through quantum teleportation, with the set-up shown in Fig. 3b. As described in the figure legend, this set-up is used to teleport the polarization state of the collective atomic excitation in a probabilistic fashion. That is, even if the protocol succeeds—that is, two of the detectors register the counts on the left-hand side—an excitation is not necessarily present in the right (target) ensembles because the product of the EME states $\rho_{L_1 R_1} \otimes \rho_{L_2 R_2}$ contains vacuum components. However, if a collective excitation appears from the right-hand side, its 'polarization' state is exactly the same as the one input from the left side. So, as in the experiment of ref. 29, such a probabilistic teleportation needs posterior confirmation of the presence of the excitation; but if the presence is confirmed, the teleportation fidelity of its polarization state is nearly perfect. The success probability for the teleportation is also given by $p_a = 1/[2(c_n + 1)^2]$, which determines the average number of repetitions needed for a final successful teleportation.

Noise and built-in entanglement purification

We now discuss noise and imperfections in our schemes for entanglement generation, connection, and applications. In particular, we show that each step contains built-in entanglement purification, which makes the whole scheme resilient to realistic noise and imperfections.

In the entanglement generation, the dominant noise is due to photon loss, which includes contributions from channel attenuation, spontaneous emissions in the atomic ensembles (which result in the population of the collective atomic mode with the accompanying photon going in other directions), coupling inefficiency of the Stokes light into and out of the channel, and inefficiency of the single-photon detectors. The loss probability is denoted by $1 - \eta_p$ with the overall efficiency $\eta_p = \eta_p' e^{-L_0/L_{\text{att}}}$, where we have separated the channel attenuation $e^{-L_0/L_{\text{att}}}$ (L_{att} is the channel attenuation length) from other noise contributions η_p' , with η_p' independent of the communication distance L_0 . The photon loss decreases the success probability for getting a detector click from p_c to $\eta_p p_c$, but it has no influence on the resulting EME state. Owing to this noise, the entanglement preparation time should be replaced by $T_0 \approx t_\Delta/(\eta_p p_c)$. The second source of noise comes from the dark counts of the single-photon detectors. The dark count gives a detector click, but without population of the collective atomic mode, so it contributes to the vacuum coefficient in the EME state. If the dark count comes up with a probability p_{dc} for the time interval t_Δ , the vacuum coefficient is given by $c_0 = p_{\text{dc}}/(\eta_p p_c)$, which is typically much smaller than 1 as the Raman transition rate is much larger than the dark count rate. The final source of noise, which influences the fidelity of getting the EME state, is caused by the event in which more than one atom is excited to the collective mode S whereas there is only one click in D_1 or D_2 . The conditional probability for that event is given by p_c , so we can estimate the fidelity imperfection $\Delta F_0 \equiv 1 - F_0$ for the entanglement

generation by:

$$\Delta F_0 \approx p_c \quad (5)$$

Note that by decreasing the excitation probability p_c , we can make the fidelity imperfection closer and closer to zero—with the price of a longer entanglement preparation time T_0 . This is the basic idea of the entanglement purification. So, in this scheme, the confirmation of the click from the single-photon detector generates and purifies entanglement at the same time.

In the entanglement swapping, the dominant noise is still due to the losses, which include contributions from detector inefficiency, the inefficiency of the excitation transfer from the collective atomic mode to the optical mode^{21,22}, and the small decay of atomic excitation during storage^{20–22}. Note that by introducing the detector inefficiency, we have automatically taken into account the imperfection that the detectors cannot distinguish between one and two photons. With all these losses, the overall efficiency in entanglement swapping is denoted by η_s . The loss in entanglement swapping gives contributions to the vacuum coefficient in the connected EME state, as in the presence of loss a single detector click might result from two collective excitations in the ensembles I_1 and I_2 , and in this case, the collective modes in the ensembles L and R have to be in a vacuum state. After taking into account the realistic noise, we can specify the success probability and the new vacuum coefficient for the i th entanglement connection by the recursion relations $p_i \equiv f_1(c_{i-1}) = \eta_s(1 - \{\eta_s/[2(c_{i-1} + 1)]\})/(c_{i-1} + 1)$ and $c_i \equiv f_2(c_{i-1}) = 2c_{i-1} + 1 - \eta_s$. The coefficient c_0 for the entanglement preparation is typically much smaller than $1 - \eta_s$, so we have $c_i \approx (2^i - 1)(1 - \eta_s) = (L_i/L_0 - 1)(1 - \eta_s)$, where L_i denotes the communication distance after i times entanglement connection. With the expression for the c_i , we can evaluate the probability p_i and the communication time T_n for establishing an EME state over the distance $L_n = 2^n L_0$. After the entanglement connection, the fidelity of the EME state also decreases, and after n times connection, the overall fidelity imperfection $\Delta F_n \approx 2^n \Delta F_0 \approx (L_n/L_0) \Delta F_0$. We need to make ΔF_n small by decreasing the excitation probability p_c in equation (5).

We note that our entanglement connection scheme also has a built-in entanglement-purification function. This can be understood as follows: each time we connect entanglement, the imperfections of the set-up decrease the entanglement fraction $1/(c_i + 1)$ in the EME state. However, this fraction decays only linearly with distance (the number of segments), which is in contrast to the exponential decay of entanglement for connection schemes without entanglement purification. The reason for the slow decay is that for each time of entanglement connection, we need to repeat the protocol until there is a detector click, and the confirmation of a click removes part of the added vacuum noise, as a larger vacuum component in the EME state results in more repetitions. The built-in entanglement purification in the connection scheme is essential for the polynomial scaling law of the communication efficiency.

As in the entanglement generation and connection schemes, our entanglement application schemes also have built-in entanglement purification, which makes them resilient to realistic noise. First, we have seen that the vacuum components in the EME states are removed from the confirmation of the detector clicks, and thus have no influence on the fidelity of all the application schemes. Second, if the single-photon detectors and the atom-to-light excitation transitions in the application schemes are imperfect, with the overall efficiency denoted by η_a , we can show that these imperfections only influence the efficiency of getting detector clicks—with the success probability replaced by $p_a = \eta_a/[2(c_n + 1)^2]$ —and have no effect on communication fidelity. Last, we have seen that the phase shifts in the stationary channels and the small asymmetry of the stationary set-up are removed automatically when we project the EME state to the PME state, and thus have no influence on the communication fidelity.

Noise not correctable by our scheme includes the detector dark count in the entanglement connection, the non-stationary channel noise and set-up asymmetries. The fidelity imperfection resulting from the dark count increases linearly with the number of segments L_n/L_0 , and the imperfections from the non-stationary channel noise and set-up asymmetries increase by the random-walk law $\sqrt{L_n/L_0}$. For each time of entanglement connection, the dark count probability is about 10^{-5} if we make a typical choice that the collective emission rate is about 10 MHz and the dark count rate is 10^2 Hz. So this noise is negligible, even if we have communicated over a long distance (10^3 times the channel attenuation length L_{att} , for instance). The non-stationary channel noise and set-up asymmetries can also be safely neglected for such a distance. For instance, it is relatively easy to control the non-stationary asymmetries in local laser operations to values below 10^{-4} with the use of accurate polarization techniques³⁰ for Zeeman sublevels (as in Fig. 2b).

Scaling of the communication efficiency

We have shown that each of our entanglement generation, connection, and application schemes has built-in entanglement purification, and as a result of this property, we can fix the communication fidelity to be nearly perfect, and at the same time require the communication time to increase only polynomially with distance. Assume that we want to communicate over a distance $L = L_n = 2^n L_0$. By fixing the overall fidelity imperfection to be a desired small value ΔF_n , the entanglement preparation time becomes $T_0 \approx t_\Delta/(\eta_p \Delta F_0) \approx (L_n/L_0)t_\Delta/(\eta_p \Delta F_n)$. For effective generation of the PME state of equation (4), the total communication time $T_{tot} \approx T_n/p_a$ with $T_n \approx T_0 \prod_{i=1}^n (1/p_i)$. So the total communication time scales with distance by the law

$$T_{tot} \approx 2(L/L_0)^2/(\eta_p p_a \Delta F_n \prod_{i=1}^n p_i) \quad (6)$$

where the success probabilities p_i, p_a for the i th entanglement connection and for the entanglement application have been specified above.

Equation (6) confirms that the communication time T_{tot} increases with distance L only polynomially. We show this explicitly by taking two limiting cases. In the first case, the inefficiency $1 - \eta_s$ for entanglement swapping is assumed to be negligibly small. We can deduce from equation (6) that in this case the communication time $T_{tot} \approx T_{con}(L/L_0)^2 e^{L_0/L_{att}}$, with the constant $T_{con} \equiv 2t_\Delta/(\eta_p' \eta_a \Delta F_n)$ being independent of the segment length and the total distance L_0 and L . The communication time T_{tot} increases with L quadratically. In the second case, we assume that the inefficiency $1 - \eta_s$ is fairly large. The communication time in this case is approximated by $T_{tot} \approx T_{con}(L/L_0)^{[\log_2(L/L_0)+1]/2+\log_2(1/\eta_s-1)+2} e^{L_0/L_{att}}$, which increases with L still polynomially (or, more accurately, sub-exponentially, but this makes no difference in practice as the factor $\log_2(L/L_0)$ is well bounded from above for any reasonably long distance). If T_{tot} increases with L/L_0 by the m th power law $(L/L_0)^m$, there is an optimal choice of segment length ($L_0 = mL_{att}$) to minimize the time T_{tot} . As a simple estimation of the improvement in communication efficiency, we assume that the total distance L is about $100L_{att}$; for a choice of the parameter $\eta_s \approx 2/3$, the communication time $T_{tot}/T_{con} \approx 10^6$ with the optimal segment length $L_0 \approx 5.7L_{att}$. This is a notable improvement over the direct communication case, where the communication time T_{tot} for getting a PME state increases with distance L by the exponential law $T_{tot} \approx T_{con} e^{L/L_{att}}$. For the same distance $L \approx 100L_{att}$, we need $T_{tot}/T_{con} \approx 10^{43}$ for direct communication, which means that for this example the present scheme is 10^{37} times more efficient.

Outlook

We have presented a scheme for implementation of quantum repeaters and long-distance quantum communication. The proposed technique allows the generation and connection of entangle-

ment, and its use in quantum teleportation, cryptography, and tests of Bell inequalities. All of the elements of the present scheme are within reach of current experimental technology, and all have the important property of built-in entanglement purification—which makes them resilient to realistic noise. As a result, the overhead required to implement the present scheme, such as the communication time, scales polynomially with the channel length. This is in marked contrast to direct communication, where an exponential overhead is required. Such efficient scaling, combined with the relative simplicity of the experimental set-up, opens up realistic prospects for quantum communication over long distances. □

Received 16 May; accepted 12 September 2001.

1. Ekert, A. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* **67**, 661–663 (1991).
2. Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **73**, 3081–3084 (1993).
3. Bennett, C. H. *et al.* Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.* **76**, 722–725 (1991).
4. Briegel, H.-J., Duer, W., Cirac, J. I. & Zoller, P. Quantum repeaters: The role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
5. Knill, E., Laflamme, R. & Zurek, W. H. Resilient quantum computation. *Science* **279**, 342–345 (1998).
6. Preskill, J. Reliable quantum computers. *Proc. R. Soc. Lond. A* **454**, 385–410 (1998).
7. Zukowski, M., Zeilinger, A., Horne, M. A. & Ekert, A. "Event-ready-detectors" Bell experiment via entanglement swapping. *Phys. Rev. Lett.* **71**, 4287–4290 (1993).
8. Cirac, J. I., Zoller, P., Kimble, H. J. & Mabuchi, H. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Phys. Rev. Lett.* **78**, 3221–3224 (1997).
9. Enk, S. J., Cirac, J. I. & Zoller, P. Photonic channels for quantum communication. *Science* **279**, 205–207 (1998).
10. Ye, J., Vernooy, D. W. & Kimble, H. J. Trapping of single atoms in cavity QED. *Phys. Rev. Lett.* **83**, 4987–4990 (1999).
11. Hood, C. J. *et al.* The atom-cavity microscope: Single atoms bound in orbit by single photons. *Science* **287**, 1447–1453 (2000).
12. Pinkse, P. W. H., Fischer, T., Maunz, T. P. & Rempe, G. Trapping an atom with single photons. *Nature* **404**, 365–368 (2000).
13. Cabrillo, C., Cirac, J. I., G-Fernandez, P. & Zoller, P. Creation of entangled states of distant atoms by interference. *Phys. Rev. A* **59**, 1025–1033 (1999).
14. Bose, S., Knight, P. L., Plenio, M. B. & Vedral, V. Proposal for teleportation of an atomic state via cavity decay. *Phys. Rev. Lett.* **83**, 5158–5161 (1999).
15. Raymer, M. G., Walmsley, I. A., Mostowski, J. & Sobolewska, B. Quantum theory of spatial and

- temporal coherence properties of stimulated Raman scattering. *Phys. Rev. A* **32**, 332–344 (1985).
16. Kuzmich, A., Mölmer, K. & Polzik, E. S. Spin squeezing in an ensemble of atoms illuminated with squeezed light. *Phys. Rev. Lett.* **79**, 481 (1998).
17. Kuzmich, A., Bigelow, N. P. & Mandel, L. Atomic quantum non-demolition measurements and squeezing. *Europhys. Lett. A* **42**, 481–486 (1998).
18. Lukin, M. D., Yelin, S. F. & Fleischhauer, M. Entanglement of atomic ensembles by trapping correlated photon states. *Phys. Rev. Lett.* **84**, 4232–4235 (2000).
19. Duan, L. M., Cirac, J. I., Zoller, P. & Polzik, E. S. Quantum communication between atomic ensembles using coherent light. *Phys. Rev. Lett.* **85**, 5643–5646 (2000).
20. Hald, J., Sørensen, J. L., Schori, C. & Polzik, E. S. Spin squeezed state: A macroscopic entangled ensemble created by light. *Phys. Rev. Lett.* **83**, 1319–1322 (1999).
21. Phillips, D. F. *et al.* Storage of light in atomic vapor. *Phys. Rev. Lett.* **86**, 783–786 (2001).
22. Liu, C., Dutton, Z., Behroozi, C. H. & Hau, L. V. Observation of coherent optical information storage in an atomic medium using halted light pulses. *Nature* **409**, 490–493 (2001).
23. Knill, E., Laflamme, R. & Milburn, G. J. A scheme for efficient quantum computation with linear optics. *Nature* **409**, 46–52 (2001).
24. Pan, J. W., Simon, C., Brukner, C. & Zeilinger, A. Feasible entanglement purification for quantum communication. *Nature* **410**, 1067–1070 (2001).
25. Roch, J.-F. *et al.* Quantum nondemolition measurements using cold trapped atoms. *Phys. Rev. Lett.* **78**, 634–637 (1997).
26. Lo, H. K. & Chau, H. F. Unconditional security of quantum key distribution over arbitrarily long distances. *Science* **283**, 2050–2056 (1999).
27. Shor, P. W. & Preskill, J. Simple proof of security of the BB84 quantum key distribution protocol. *Phys. Rev. Lett.* **85**, 441–444 (2000).
28. Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
29. Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
30. Budker, D., Yashuk, V. & Zolotarev, M. Nonlinear magneto-optic effects with ultranarrow width. *Phys. Rev. Lett.* **81**, 5788–5791 (1998).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

This work was supported by the Austrian Science Foundation, the Europe Union project EQUIP, the ESF, the European TMR network Quantum Information, and the NSF through a grant to the ITAMP and ITR program. L.-M.D. was also supported by the Chinese Science Foundation.

Correspondence and requests for materials should be addressed to J.I.C. (e-mail: ignacio.cirac@uibk.ac.at).

Primitive deuterostomes from the Chengjiang Lagerstätte (Lower Cambrian, China)

D.-G. Shu*, S. Conway Morris†, J. Han*, L. Chen*, X.-L. Zhang*, Z.-F. Zhang*, H.-Q. Liu*, Y. Li* & J.-N. Liu*

* Early Life Institute and Department of Geology, Northwest University, Xi'an, China

† Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

Cambrian fossil-Lagerstätten (sites of exceptional fossil preservation), such as those from Chengjiang (Lower Cambrian) and the Burgess Shale (Middle Cambrian), provide our best window into the Cambrian 'explosion'. Such faunas are known from about 40 localities, and have yielded a widely disparate series of taxa ranging from ctenophores to agnathan fish. Recent excavations of the Chengjiang fossil-Lagerstätte, known from a series of sites near Kunming in Yunnan, south China, have resulted in the discovery of several new forms. In conjunction with material described earlier, these provide evidence for a new group of metazoans, the vetulicolians. Several features, notably a series of gill slits, suggest that this group can throw light on an early stage of deuterostome diversification.

The origin and evolution of metazoan bodyplans, an important aspect of the Cambrian 'explosion', represents a productive area for interdisciplinary research, between molecular biology and palaeontology^{1,2}. The extant representatives of the various phyla are, however, typically disparate in anatomy, and may be markedly different from their Cambrian ancestors. As a result, molecular phylogenies do not reveal how different bodyplans emerged, what the animals actually looked like, or how they functioned and what ecological milieux they inhabited. This historical dimension may be revealed by the fossil record, where potentially the assembly of character states can be traced within a given stem group³.

Deuterostome form is remarkably protean. Few characters are widely shared, and apart from embryological features, such as the fate of the blastopore, which does not form the mouth and can be the presumptive anus (to which exceptions are known⁴), and radial cleavage, the most notable macroscopic features in common are the gill slits^{5,6} and endostyle⁷ (albeit in terms of the dorsal epibranchial ridge in the primitive enteropneust hemichordates⁸). Gill slits, however, have been lost in all extant echinoderms, and neither is there an unequivocal trace of the endostyle in this phylum. In other cases, such as the purported homology between the hemichordate stomochord and chordate/urochordate notochord, doubt remains as to the validity of such a comparison^{8,9}. Apart from recognition of certain striking larval similarities, notably between the tornaria larvae of enteropneusts and certain echinoderm larvae, overall classical methods of comparison, such as those largely based on embryology¹⁰, led to relatively little progress until rejuvenation with molecular methods. An emerging consensus^{11–16} points to a series of significant similarities between the chordates (vertebrates and the more primitive cephalochordates) and urochordates. Evidence is also growing for a closer relationship between the echinoderms and hemichordates^{14–16}.

Even so, the evolution and ecology of the first deuterostomes remains highly conjectural. Here we report a suite of fossils, representing at least four taxa, from the Chengjiang fossil-Lagerstätte (south China). They are characterized by a bipartite body, the anterior section of which has a series of perforations, evidently precursors of gill slits. This clade is referred to as the vetulicolians, which as the class Vetulicolida was erected by Chen and Zhou¹⁷ to encompass the genera *Vetulicola* and *Banffia*. Here we expand the concept to include *Xidazoon*¹⁸, *Didazoon* nov. gen., and

*Pomatrum*¹⁹, and define a distinctive bodyplan (equivalent to a phylum) designated the Vetulicolia. These animals appear to be primitive deuterostomes.

Phylum Vetulicolia nov.

Class Vetulicolida Chen & Zhou 1997¹⁷

Family Didazoonidae Shu & Han fam. nov.

Genus *Didazoon* Shu & Han gen. nov.

Didazoon haoae Shu & Han sp. nov.

Etymology. The generic name abbreviates in Chinese the China University of Geosciences. The species name is to honour Y. Hao, for her encouragement of this work.

Holotype. Early Life Institute, Northwest University, Xi'an, ELI-0000196.

Referred material. ELI 0000197–0000217.

Stratigraphy and locality. Qiongzhusi (Chiungchussu) Formation, Yu'an-shan Member (*Eoredlichia* zone), Lower Cambrian. Specimens collected in Dabanqiao area (Kunming), about 60 km northwest of Chengjiang, and 50 km northeast of Haikou, from where *Xidazoon* was collected.

Diagnosis. Bipartite, cuticularized body, anterior segmented region and voluminous mouth, ventral margin flattened, widens posteriorly. On either side the anterior bears five circular structures, in the form of a cowl with posteriorly directed opening and basin-like interior, apparently connected to interior. A prominent constriction separates dorsal region of anterior from posterior section. The latter, composed of seven segments, tapers in both directions, with rounded posterior termination. Internal anatomy includes alimentary canal, possibly voluminous in anterior, and in posterior narrow intestine, straight or occasionally coiled. Dark strand located along ventral side of anterior section, possibly representing endostyle.

The body of *Didazoon haoae* is composed of two sections (Fig. 1a, c, d, f). On the assumption that *Didazoon* is related to *Vetulicola* (see below), where the segmented 'tail' (that is, the posterior section) evidently arises from the dorsal side, this would indicate a comparable orientation in *Didazoon*. The anterior unit is semi-quadrate in lateral view, and was laterally compressed. The leading boundary of the anterior section shows a weak pleating and is arcuate (Fig. 1a, b, d, e), suggesting a voluminous opening. The anterior section is now filled with sediment, an arrangement indicating that in life the anterior cavity was spacious. The anterior rim and dorsoventral

margins, however, are flatter. In particular the ventral margin has a bladed appearance (Fig. 1a, b, d, e), with a progressive widening posteriorly. The exterior shows a prominent series of segments, totalling six. The membranes separating the segments are fairly broad, and this may have conferred a degree of flexibility.

Laterally, five oval-shaped structures are located on either side of the anterior section (Fig. 1a–f). They are interpreted as gills. The anteriormost is smallest, the next two are the largest and may imbricate, and the two posteriormost show a progressive size decrease. Each consists of a convex cowl, sometimes with concentric lineations, with a posterior opening. An internal cavity has a basin-like floor. An opening into the interior is inferred to exist towards the anterior of each cavity. Faint imprints of apparently tubular units may represent internal connections between these structures. Towards the ventral margin there is a darker strand of material (Fig. 1a, b, d, e), possibly representing the endostyle, as it is a consistent feature, which extends posteriorly and further back rises dorsally towards the narrow alimentary canal that extends into the posterior section.

The constriction shows creasing and was probably quite flexible (Fig. 1a, d), suggesting a degree of independence between anterior and posterior regions. The elongate posterior section tapers in either direction. It consists of seven segments, of which the last has a rounded termination. A narrow gut trace is well preserved, and sometimes filled with fine-grained material. In one specimen the infill has a spiral arrangement, which is also seen in *Vetulichola* (Fig. 5a, b and Supplementary Information Fig. III). Towards either margin narrow strands, sometimes branched, may represent vascular tissue.

New information on *Xidazoon* and *Vetulichola*

Continued collecting has led to new data on *Xidazoon*¹⁸ and *Vetulichola*^{17,20,21}. Since the original description¹⁸ of *Xidazoon* more

specimens (Fig. 2a–f and Supplementary Information Fig. I) have been found from Mafang and Dazicun, in the Haikou area²². These show some important but hitherto unrecognized features. In particular, the anterior section bears a series of up to five cowl-like structures (Figs 2a, b, d, e and 3c, d), similar to those of *Didazoon* in shape, size and location. In one specimen, associated dark, circular areas are possibly indicative of the actual openings into the interior (Fig. 2d). A step-like arrangement along the right-hand margin of the holotype¹⁸ is probably the hitherto unrecognized lateral expression of these openings. The prominent circular feeding apparatus is well preserved in several specimens (Fig. 2c, f and Supplementary Information Fig. I). Despite distortion it shows approximately 30 plates, thus exceeding the number estimated to occur in the holotype¹⁸. A persistent feature is a rather dark region running close to the ventral and posterior margins (Fig. 2a, d), which is quite similar to that in *Didazoon*, and is tentatively interpreted as an endostyle. Folding in the anterior section (Fig. 2c and Supplementary Information Fig. I) suggests that in life it was thin-walled, although one specimen shows that the presumed venter had a keel-like arrangement (Supplementary Information Fig. I).

Didazoon and *Xidazoon*, which are found in localities separated by about 50 km and may belong to different faunal assemblages, share many similarities, but the differences justify generic separation. *Didazoon* has a more cuticularized anterior section with a series of clear segmental boundaries, whereas in *Xidazoon* only the first segment is moderately well defined. The greater flexibility of the anterior section of *Xidazoon* is reflected in both its folding and a propensity for the more posterior region to decay. The anterior margin of *Didazoon*, similar to that of *Vetulichola rectangulata* (Fig. 4f and Supplementary Information Fig. I), appears to have been more rigid and may have had a less flexible feeding apparatus. The constriction in this taxon appears to be better defined and shorter

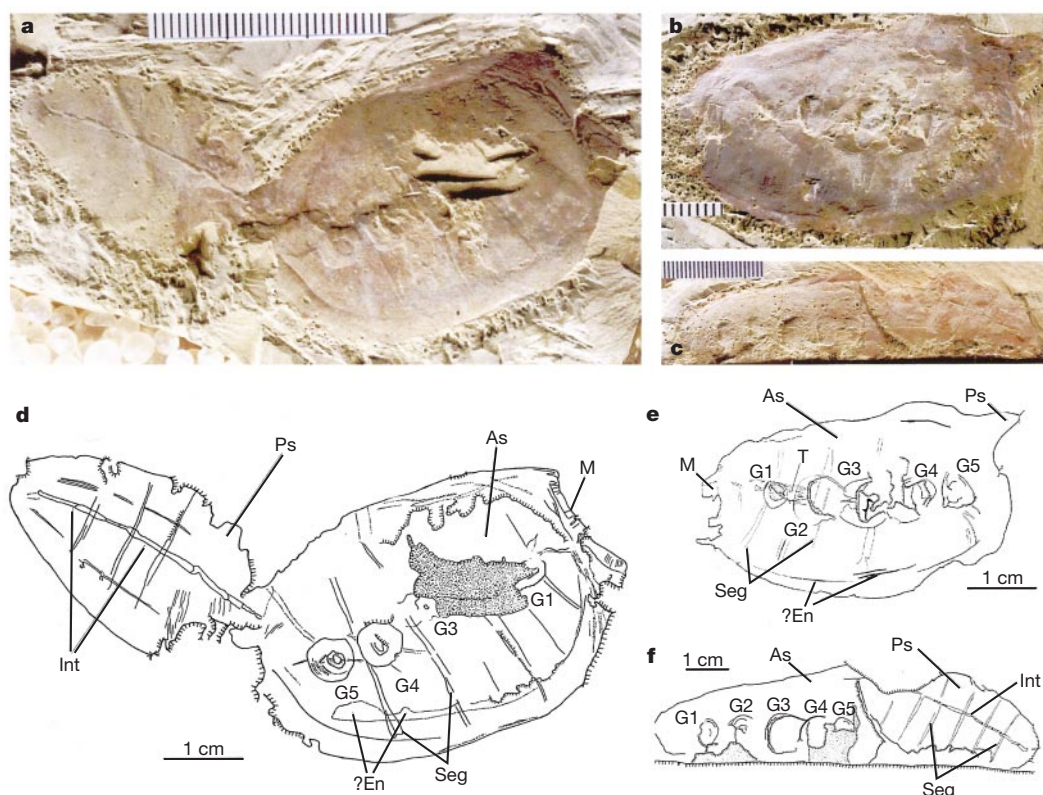


Figure 1 The Lower Cambrian vetulicolian *Didazoon haoae* Shu & Han gen. et sp. nov. from Dabanqiao, Yunnan. **a**, Specimen ELI-0000196, entire specimen, compare **d**. **b**, ELI-0000198, anterior section and remnant of posterior, compare **e**. **c**, ELI-0000200A,

anterior incomplete but with prominent gills, segmented posterior with intestine, compare **f**. In **a–c**, millimetric scale bars are shown. As, anterior section; ?En, presumed endostyle; G1–5, gill1–gill5; Int, intestine; M, mouth; Ps, posterior section; Seg, segments.

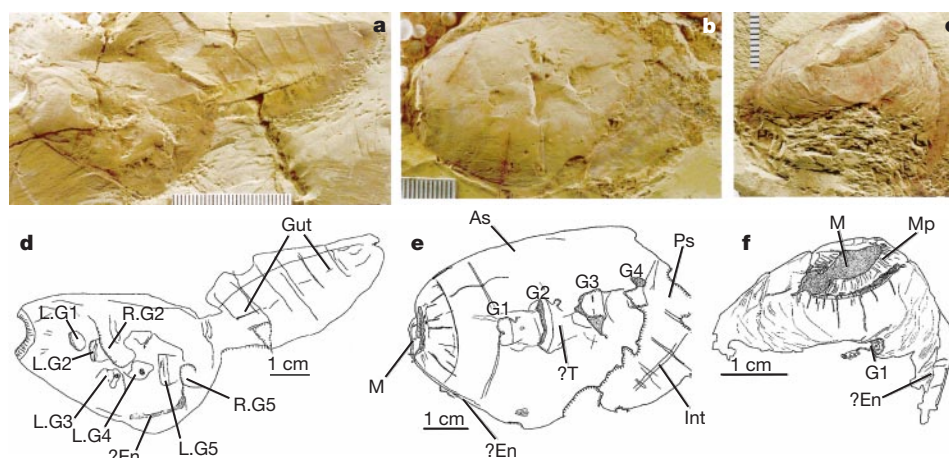


Figure 2 The Lower Cambrian vetulicolian *Xidazoon stephanus* from Haikou, Kunming. **a**, ELI-0000203A, entire specimen, although the anterior is crushed it appears to show five left gills and two right gills, including possible openings, compare **d**. **b**, ELI-0000202, entire specimen with gills, compare **e**. **c**, ELI-0000204A, anterior section, incomplete but

with well preserved mouth, compare **f**. In **a–c**, millimetric scale bars are shown. As, anterior section; ?En, presumed endostyle; L. G1–5, left gill1–gill5; R. G2, 5, right gills 2, 5; Int, intestine; M, mouth; Mp, mouth plates; Ps, posterior section; ?T, possible tube connecting gills.

than in *Xidazoon*. Moreover, they differ in the gill structures (compare Fig. 1a, d with Fig. 2a, b, d, e).

Vetulicola is represented by two species, *V. cuneata*^{17,20,21} (Fig. 4a–e and Supplementary Information Fig. I) and *V. rectangulata*¹⁹ (Fig. 4f, g and Supplementary Information Fig. I). Their principal difference is in the anterior: the former species shows prominent lobate extensions, whereas in *V. rectangulata* the anterior margin is almost straight in lateral view, similar to that in *Didazoon*. *Vetulicola* has a strongly bipartite bodyplan (Fig. 4a), reminiscent of phyllocarid arthropods. The laterally compressed anterior section forms a sort of ‘carapace’, consisting of four prominent and rather rigid plates. A sediment infill indicates the interior was voluminous. The venter forms a keel-like structure (Fig. 4g), and the lateral midlines consist of a narrow groove associated with five openings (Fig. 4a–f and Supplementary Information Fig. II). At its posterior end there is a median and arcuate dorsal process (Fig. 5f). Little is known about the anatomy of the interior. Excavation has failed to reveal evidence for limbs or other appendages within the carapace-like structure, nor are there signs of eyes. Their inclusion in reconstructions (in Fig. 137 of ref. 17; p. 38 of ref. 21) seems to be questionable. A series of small, blunt pegs lining the interior, which give the anterior section a characteristically ‘beaded’ appearance, may have been used to attach tissue, possibly involved with food collection.

The most obvious feature of the anterior section is a series of openings¹⁷, located along the lateral midlines. They total five on each side, and have a complex structure. Each consists of a prominent and spacious pouch. This may be filled with sediment, but when empty each shows a low ridge with filament-like structures on the floor (Fig. 3e–l and Supplementary Information Fig. II). Anterior to this is a steep wall that near its base has an elongate area. Typically this is filled with sediment, and is interpreted as the inhalant aperture. A peculiar feature of the internal cavity is that it extends posteriorly to define a prominent tapering tube sloping upwards towards the outer surface. It is not clear whether this tube opened into the next-posterior pouch, in a position above its inhalant aperture, or was blind. The exhalant aperture is located in the lateral groove, and consists of an elliptical opening surrounded by a series of very fine filaments (Figs 3e–j and 4c; see also Supplementary Information Fig. II). Adjacent to the exhalant openings, the cuticle on the dorsal and ventral plates forms semi-elliptical structures (Fig. 3i, j and Fig. 4a, f). These seem previously¹⁷ to have been mistaken for the actual openings, but as they lie above the apertures they evidently represent protective

cuticular thickenings. Each complex is interpreted as a gill, with internal respiratory filaments.

The posterior region forms a tail (Fig. 4a, e and Supplementary Information Fig. I) consisting of seven somites, not eight as reported¹⁷, each strongly cuticularized, and separated by broad membranes. The last segment has a rounded termination. The tail region is dorsoventrally asymmetric, and probably provided propulsive strokes by lateral movements. The posterior region often demonstrates a gut trace, sometimes coiled and with prominent fine-grained contents (Fig. 5b and Supplementary Information III).

The *Xidazoon*–*Didazoon* connection is further extended by reference to the already published, but poorly known, Chengjiang taxon *Pomatrum ventralis*¹⁹. The anterior section and feeding circlet is similar to those in *Didazoon*. On the anterior of an incomplete specimen there are two quadrate cowls, the forward conspicuously smaller in size, evidently equivalent to the gills of *Didazoon*. The major difference between *Pomatrum* and *Didazoon* (as well as *Xidazoon*) concerns the posterior section, which extends from the dorsal side as a narrow tail-like structure, as in *Vetulicola*.

The vetulicolian clade (*Didazoon*, *Pomatrum*, *Vetulicola* and *Xidazoon*) is well defined by the following diagnostic features: a bipartite body, consisting of a voluminous anterior with five sets of gill openings; and a posterior section with seven segments. These taxa appear to occupy various ecologies. *Didazoon* and *Xidazoon* were probably benthic and epifaunal, whereas the box-like anterior and elongate tail of *Vetulicola* indicate greater mobility, perhaps as an active swimmer. *Vetulicola* is rather arthropod-like, but we interpret this as an example of evolutionary convergence. A carapace construction consisting of four large plates is unknown in arthropods. Most importantly, there is no evidence for any sort of associated appendage. In addition, there is no evidence that *Vetulicola* moulted its skeleton. Growth was presumably by accommodation, and the presence of surface membranes around all the margins and along the lateral mid-lines of the animal (Fig. 5c–f and Supplementary Information Fig. IV) suggests that the skeleton was internal.

The most striking difference between the vetulicolians and arthropods, however, concerns the series of five gills. In *Xidazoon*, *Didazoon* and *Pomatrum*, they consist of hood-like structures that in life concealed the openings. In *Vetulicola* these gills are more complex, but their homology is in little doubt both because of the overall similarities of the bodyplan and *Didazoon* and *Pomatrum* acting as a morphological intermediate between *Xidazoon* and

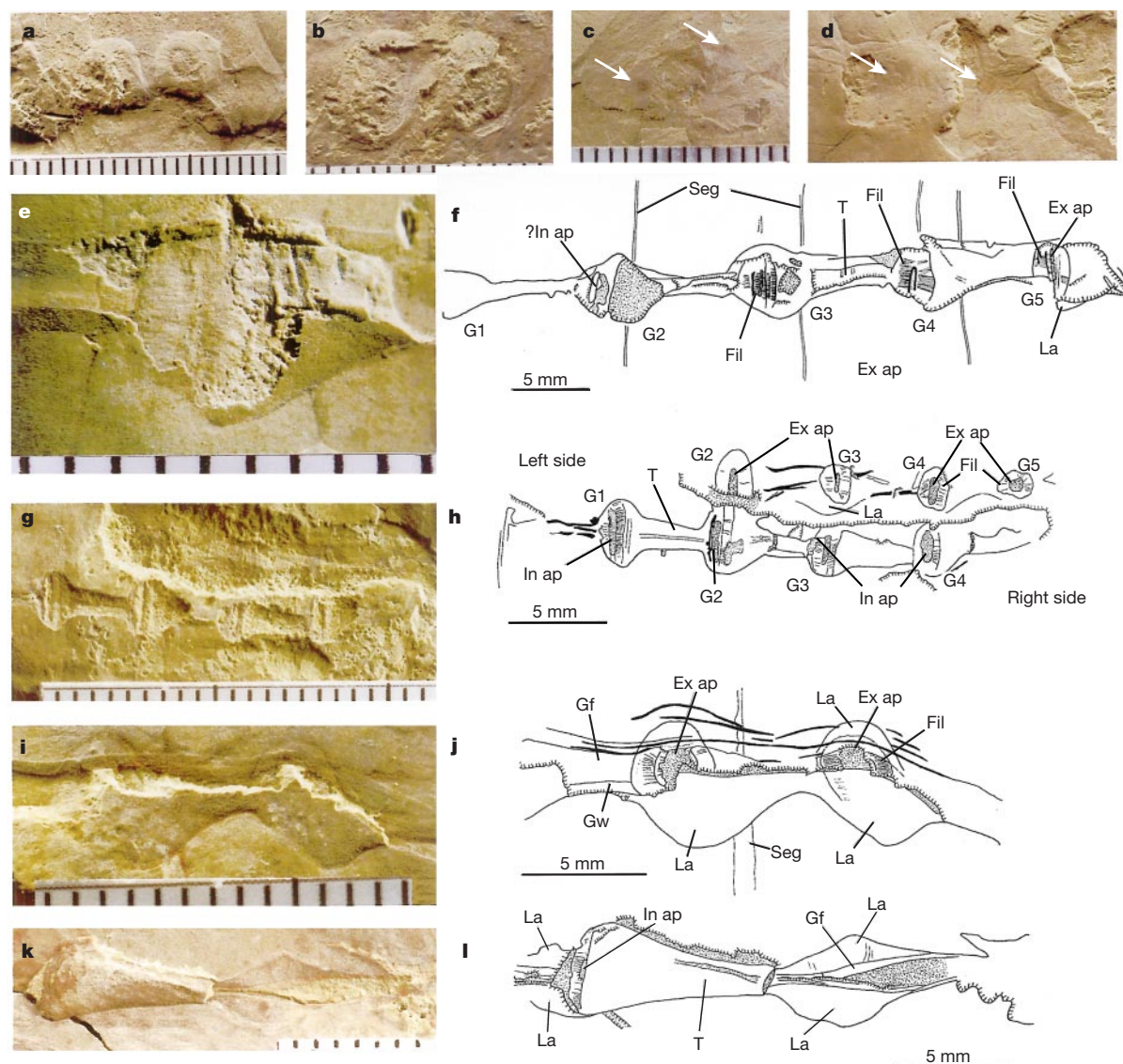


Figure 3 Gill structures of vetulicolians. **a–d**, Details of gills of specimens ELI-0000196 (*Didazoon*), 201 (*Didazoon*), 203A (*Xidazoon*) and 202 (*Xidazoon*), respectively, compare Figs 1a and 2a. **e, f**, Details of the left gill 4 and all left gills of ELI-0000216 (*Vetulicola cuneata*), respectively; compare Fig. 4a. **g, h**, Details of gills on two levels representing respectively lower (left side, gills G2–5) and upper (right side, gills G1–4) of

ELI-0000215 (*Vetulicola cuneata*), compare Fig. 4e. **i, j**, Details of left gills 4 and 5 of ELI-0000210 (*Vetulicola rectangulata*). **k, l**, Details of gills viewed from interior of specimen ELI-0000212 (*Vetulicola rectangulata*). In **a–e, g, i, k**, millimetric scale bars are shown. Ex ap, exhalant aperture; Fil, filaments; Gf, groove floor; Gw, groove wall; G1–5, gill1–gill5; In ap, inhalant aperture; La, lappet; Seg, segments; T, tube.



Figure 4 The Lower Cambrian vetulicolian *Vetulicola cuneata* (**a–e**) from Chengjiang and *Vetulicola rectangulata* (**f, g**) from Haikou, Kunming. **a**, ELI-0000216, compare Fig. 3e. **b–d**, Specimen ELI-0000207, two animals, upper with interior excavated to reveal right-hand gill pouches (**c**), lower with well-preserved gills, detail shown in **d**.

e, ELI-0000215, entire specimen and detail of gills on left- and right-hand sides of the anterior section, compare Fig. 3g. **f**, ELI-0000209, anterior section with lappets on gills. **g**, ELI-0000211A, ventral view of anterior section showing mouth cone. Millimetric scale bars are shown.

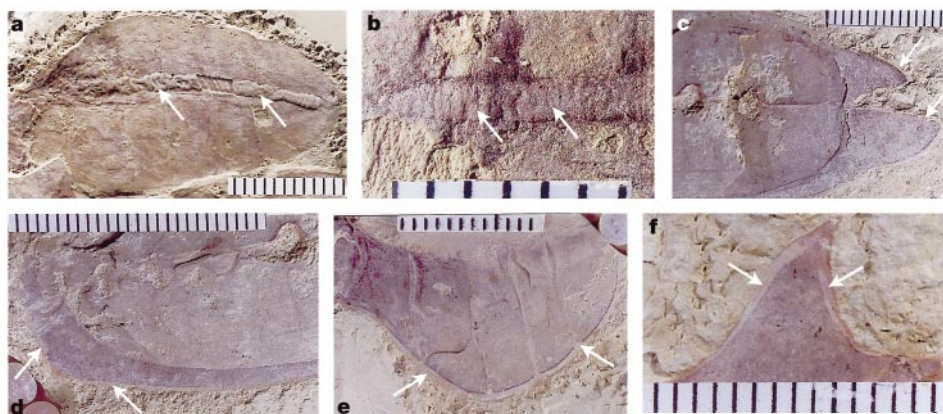


Figure 5 Coiled guts (arrows in **a** and **b**) and surface membranes (arrows in **c–f**) of vetulicolians. **a**, ELI-0000197 (*Didazoon*). **b**, ELI-0000309A (*Vetulicola*). **c–f**, *Vetulicola*. **c**, Anterior of ELI-0000218. **d**, Postero-ventral of the anterior section in ELI-0000218.

e, Posterior section of ELI-0000302B. **f**, Dorsal ‘fin’ of ELI-0000338A. Millimetric scale bars are shown.

Vetulicola. In *Xidazoon* and *Didazoon* the gills are relatively simple, and filamentous structures comparable to those in *Vetulicola* have not been observed. The pouch-like gills of *Vetulicola* are probably a reflection of both a more active mode of life and more rigid construction of the body, possibly decreasing overall gas permeability.

Vetulicolians as deuterostomes

Gill slits are, of course, one of the defining features of the deuterostomes. In the relative simple form, as seen in *Xidazoon* and *Didazoon*, their primary function may have been discharge of excess sea water. As such, they would have achieved what is believed to have been the original function of the deuterostome gill slits²³, although functionally they are usually also used for respiration and/or suspension feeding. We acknowledge that given the obvious dissimilarity of these apertures to the gill slits of even the most primitive of extant deuterostomes, widely regarded as the hemichordate acorn worms, it might be more sensible to interpret the lateral openings of vetulicolians as an example of evolutionary convergence. The gill slits of the acorn-worms are, however, evidently specialized, to judge from their striking similarity to the equivalent structures in amphioxus, although within the deuterostomes the hemichordates and cephalochordates are phylogenetically relatively remote^{14–16}. If the original function of the apertures was indeed the disposal of sea water^{23,24}, the familiar close packing of gill slits seen in most deuterostomes would have resulted principally from the demands of respiration. Moreover, the gill slits (although their interpretation is controversial) of cornute calcichordates²⁵, such as *Cothurnocystis*, are also large and widely spaced, and it is only in derived forms, such as *Scotiaecystis*, that these gill slits again become more closely packed.

Two other features in the vetulicolians may be consistent with a position within the deuterostomes. One is the dark ventral strand in

the anterior section, possibly equivalent to an endostyle. Second, the evidence that the skeleton was internal suggests that it was of mesodermal derivation. Further evidence for a deuterostome affinity may, however, lie with a proposed link between the vetulicolians and the Chengjiang clade represented by *Haikouella*²⁶ and *Yunnanozoon*^{27–29}. These animals were interpreted as craniates, but while few dispute features such as the gill slits, many of the interpretations of a specifically craniate anatomy³⁰ are open to question. It seems likely that these taxa are more basal within the deuterostomes, with a position closer to the hemichordates^{28,31}. Despite obvious differences, vetulicolians and yunnanozoans may be related. The former has a bipartite bodyplan, while yunnanozoans are divided into an anterior section (with proboscis and collar)²⁸, bearing a series of gill slits, and a segmented and cuticular posterior. The yunnanozoan body is elongate and vaguely fish-like, but evidence for partial separation between the two regions (Fig. 6 and Figs 2e and 3d in ref. 28) suggests that the yunnanozoan bodyplan arose from of an overlap of two discrete sections.

Discussion

Vetulicola was widely accepted as “a very strange arthropod”^{17,19,32,33} (but see ref. 34). Structures that perforate the anterior body wall are, however, interpreted as the precursors of the gill slits³⁴, a defining character of the deuterostomes. Similarities between the vetulicolians and yunnanozoans may give an important clue as to a subsequent stage in deuterostome evolution, given that the gill slits of the latter are closer in organization to other primitive deuterostomes. The bipartite body of vetulicolians is also reminiscent of the hypothetical animal invoked by Romer’s somato-visceral theory³⁵. This posited an effectively bipartite organization in which the ‘somatic’ component (with musculature and a segmented nervous system) became increasingly integrated with a visceral unit that was largely composed of a pharyngeal branchial basket.

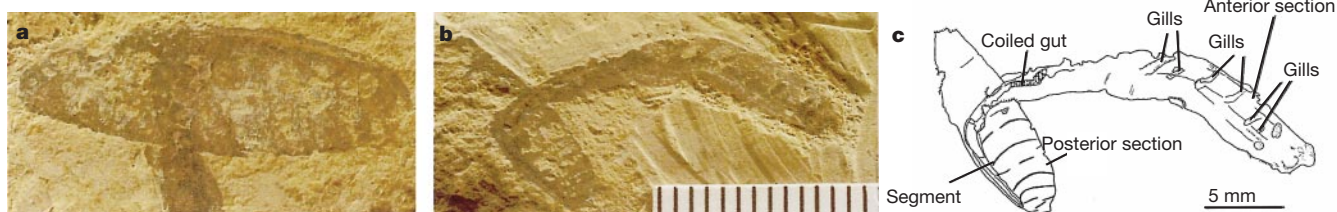


Figure 6 The Lower Cambrian *Yunnanozoon lividum* from Chengjiang, Yunnan. **a**, Part, complete specimen, Specimen NWU93-1406A. **b**, Counterpart, detail of posterior section

and its attachment to anterior section, NWU93-1406B. **c**, Camera-lucida drawing with details of part and counter-part combined. In **b**, a millimetric scale bar is shown.

Although originally applied to the first chordates, this concept might be more relevant to the appearance of basal deuterostomes. In addition, the coiled gut commonly seen in both vetulicolians (Fig. 5a, b and Supplementary Information Fig. III) and *Yunnanozoon* (Fig. 6 and Figs 2e and 3d in ref. 28), as well as in various less-advanced fish³⁶, is also consistent with deuterostome affinities of vetulicolians. An alternative although less likely possibility is that this coiled structure incorporates the equivalent of the notochord, perhaps as a series of stacked discs. The sporadic occurrence of the coiling and irregular convolutions, however, make a gut interpretation more probable (see Supplementary Information Figs I and III).

Antecedents of vetulicolians are problematic. Possibly the most primitive representative of this clade is *Banffia*, known from the Burgess Shale³⁷ and Chengjiang^{17,21}. A detailed description is not yet available, but the correspondences include the bipartite body connected by a narrow constriction, and a segmented posterior, albeit with finer divisions than other vetulicolians. The anterior bore a mouth with a circlet of cuticular folds (J. Caron, personal communication), but equivalents to the vetulicolian gills have not been recognized.

If vetulicolians represent primitive deuterostomes, their body-plan could point to a deeper origination, somewhere within the protostomes. To date, however, neither the Neoproterozoic nor Cambrian fossil record seems to contain taxa that might serve as such an evolutionary link. The role of vetulicolians in the early deuterostome evolution is, however, more intriguing. The extant phyla are both highly disparate and derived. Various differences in gene expression patterns, especially between the hemichordates and more advanced chordates, indicate redeployments³⁸, such as the use of *Brachyury*^{9,39}, or significant re-organization as in the nervous system^{40–42}. In contrast, the gill slits^{5,6} and probably the endostyle⁸ are primitive features. It is worth mentioning that *Vetulicola* itself has a certain resemblance to a giant 'tadpole' (Fig. 4a). This concept has played a central role in speculations on the origins of the chordates⁴³. The development of a propulsive tail and a large anterior feeding chamber with pharyngeal openings may in itself give further insights into the subsequent evolution of the chordate bodyplan, including the calcichordates. Finally, we note that the co-occurrence of stem-group deuterostomes and agnathan fish⁴⁴ are consistent with an 'explosion' of metazoan bodyplans in the latest Neoproterozoic and early Cambrian. □

Received 1 June; accepted 13 September 2001.

- Conway Morris, S. Evolution: Bringing molecules into the fold. *Cell* **100**, 1–11 (2000).
- Valentine, J. W., Jablonski, D. & Erwin, D. H. Fossils, molecules and embryos: new perspectives on the Cambrian explosion. *Development* **126**, 851–859 (1999).
- Budd, G. E. & Jensen, S. A critical reappraisal of the fossil record of the bilaterian phyla. *Biol. Rev.* **75**, 253–295 (2000).
- Fioroni, P. Zur Signifikanz des Blastoporus-Verhaltens in evolutiver Hinsicht. *Rev. Suisse Zool.* **87**, 261–272 (1980).
- Ogasawara, M. *et al.* Developmental expression of *Pax 1/9* genes in urochordate and hemichordate gills: insight into function and evolution of the pharyngeal epithelium. *Development* **125**, 2539–2550 (1999).
- Okai, N. *et al.* Characterization of gill-specific genes of the acorn worm *Ptychodera flava*. *Dev. Dyn.* **217**, 309–319 (2000).
- Ogasawara, M. *et al.* Ascidian homologs of mammalian thyroid transcription factor-1 gene are expressed in the endostyle. *Zool. Sci.* **16**, 559–565 (1999).
- Ruppert, E. E., Cameron, C. B. & Frick, J. F. Endostyle-like features of the dorsal epibranchial ridge of an enteropneust and the hypothesis of dorsal-ventral axis inversion in chordates. *Invert. Biol.* **118**, 202–212 (1999).
- Peterson, K. J. *et al.* A comparative molecular approach to mesodermal patterning in basal deuterostomes: the expression pattern of *Brachyury* in the enteropneust hemichordate *Ptychodera flava*. *Development* **126**, 85–95 (1999).
- Schaeffer, B. Deuterostome monophyly and phylogeny. *Evol. Biol.* **21**, 179–235 (1987).

- Turbeville, J. M., Schulz, J. R. & Raff, R. A. Deuterostome phylogeny and the sister group to the chordates: evidence from molecules and morphology. *Mol. Biol. Evol.* **11**, 648–655 (1994).
- Wada, H. & Satoh, N. Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18S rDNA. *Proc. Natl Acad. Sci. USA* **91**, 1801–1804 (1994).
- Lucalli, T. C. The nature and origin of deuterostomes: some unresolved issues. *Invert. Biol.* **116**, 363–370 (1997).
- Bromham, L. D. & Degnan, B. M. Hemichordate and deuterostome evolution: robust molecular phylogenetic support for a hemichordate + echinoderm clade. *Evol. Dev.* **1**, 166–171 (1999).
- Cameron, C. B., Garey, J. R. & Swalla, B. J. Evolution of the chordate body plan: New insights from phylogenetic analyses of deuterostome phyla. *Proc. Natl Acad. Sci. USA* **97**, 4469–4474 (2000).
- Gee, H. in *Major Events in Early Vertebrate Evolution* (ed. Ahlberg, P. E.) *Syst. Ass. Spec.* **61**, 1–14 (2001).
- Chen, J.-Y. & Zhou, G.-Q. Biology of the Chengjiang fauna. *Bull. Natl Mus. Nat. Sci. Taiwan* **10**, 11–105 (1997).
- Shu, D.-G. *et al.* A pipiscid-like fossil from the Lower Cambrian of south China. *Nature* **400**, 746–749 (1999).
- Luo, H.-L. *et al.* Early Cambrian Chengjiang Fauna from Kunming Region, China (Yunnan Sci. Technol. Press, Kunming, 1999).
- Hou, X.-G. Early Cambrian large bivalved arthropods from Chengjiang, eastern Yunnan. *Acta Palaeont. Sinica* **26**, 286–297 (1987).
- Chen, J.-Y. *et al.* The Chengjiang Biota: A Unique Window of the Cambrian Explosion (National Museum of Natural Science, Taiwan, 1996).
- Zhang, X.-L. *et al.* New sites of Chengjiang fossils: crucial windows on the Cambrian explosion. *J. Geol. Soc. Lond.* **158**, 211–218 (2001).
- Gilmour, T. H. J. Feeding in pterobranch hemichordates and the evolution of gill slits. *Can. J. Zool.* **57**, 1136–1142 (1979).
- Gilmour, T. H. J. Feeding in tornaria larvae and the development of gill slits in enteropneust hemichordates. *Can. J. Zool.* **60**, 3010–3020 (1982).
- Jefferies, R. P. S. *The Ancestry of the Vertebrates* (British Museum (Natural History), London, 1986).
- Chen, J.-Y., Huang, D. Y. & Li, C. W. An early Cambrian craniate-like chordate. *Nature* **402**, 518–522 (1999).
- Chen, J.-Y. *et al.* A possible early Cambrian chordate. *Nature* **377**, 720–722 (1995).
- Shu, D.-G., Zhang, X. & Chen, L. Reinterpretation of *Yunnanozoon* as the earliest known hemichordate. *Nature* **380**, 428–430 (1996).
- Dzik, J. *Yunnanozoon* and the ancestry of chordates. *Acta Palaeont. Polonica* **40**, 341–360 (1995).
- Holland, N. D. & Chen, J.-Y. Origin and early evolution of the vertebrates: new insights from advances in molecular biology, anatomy, and paleontology. *BioEssays* **23**, 142–151 (2001).
- Shu, D.-G., Chen, L., Zhang, X.-L., Han, J. & Li, Y. Chengjiang Lagerstätte and earliest-known chordates. *Zool. Sci.* **18**, 447–448 (2001).
- Hou, X.-G. *et al.* The Chengjiang Fauna: Exceptionally Well-preserved Animals From 530 Million Years Ago (Yunnan Sci. Technol. Press, Kunming, 1999).
- Hou, X.-G. & Bergström, J. Arthropods of the Lower Cambrian Chengjiang fauna, southwest China. *Fossils Strata* **45**, 1–116 (1997).
- Shu, D.-G., Zhang, X.-L. & Chen, L. in *Progress in Geology of China (1993–1996)* (Papers to 30th International Geological Congress) 42–45 (Chinese Geological Society, Beijing, 1996).
- Romer, A. S. The vertebrate as a dual animal—somatic and visceral. *Evol. Biol.* **6**, 121–156 (1972).
- Romer, A. S. *The Vertebrate Body* (Saunders, Philadelphia, 1964).
- Walcott, C. D. Middle Cambrian annelids. *Smithson. Misc. Coll.* **57**, 109–144 (1911).
- Taguchi, S. *et al.* Characterization of a hemichordate fork head/HNF-3 gene expression. *Dev. Genes Evol.* **210**, 11–17 (2000).
- Tagawa, K., Humphreys, T. & Satoh, N. Novel pattern of *Brachyury* gene expression in hemichordate embryos. *Mech. Dev.* **75**, 139–143 (1998).
- Lucalli, T. C. Apical organs, epithelial domains, and the origin of the chordate central nervous system. *Am. Zool.* **34**, 533–541 (1994).
- Tagawa, K., Humphreys, T. & Satoh, N. *T-Brain* expression in the apical organ of hemichordate tornaria larvae suggests its evolutionary link to the vertebrate forebrain. *J. Exp. Zool. (Mol. Dev. Evol.)* **288**, 23–31 (2000).
- Nielsen, C. Origin of the chordate central nervous system and the origin of chordates. *Dev. Genes Evol.* **209**, 198–205 (1999).
- Gee, H. *Before the Backbone: Views on the Origin of the Vertebrates* (Chapman & Hall, London, 1996).
- Shu, D.-G. *et al.* Lower Cambrian vertebrates from South China. *Nature* **402**, 42–46 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

Supported by Ministry of Sciences and Technology of China, Natural Science Foundation of China, Education Department of China, and National Geographic, USA (D.-G.S., J.H., L.C., X.-L.Z., Z.-F.Z., H.-Q.L., Y.L. and J.-N.L.), the Royal Society, and St. John's College, Cambridge (S.C.M.). H.-L. Luo and S.-X. Hu gave access to material of *Pomatum* in Kunming. We thank R. J. Aldridge, P. Janvier and R. P. S. Jefferies for comments, and L.-H. Guo, X.-X. Cheng, H.-X. Guo, S. J. Last and S. Capon for technical assistance.

Correspondence and requests for materials should be addressed to D.-G.S. (e-mail: dgshu@sein.xgbc.com.cn).

Regulation of the X-ray luminosity of clusters of galaxies by cooling and supernova feedback

G. Mark Voit* & Greg L. Bryan†‡

* Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA

† Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

‡ Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK

Clusters of galaxies are thought to contain about ten times as much dark matter as baryonic matter¹. The dark component therefore dominates the gravitational potential of a cluster, and the baryons confined by this potential radiate X-rays with a luminosity that depends mainly on the gas density in the cluster's core². Predictions of the X-rays' properties based on models of cluster formation do not, however, agree with the observations. If the models ignore the condensation of cooling gas into stars and feedback from the associated supernovae, they overestimate the X-ray luminosity because the density of the core gas is too high. An early episode of uniformly distributed supernova feedback could rectify this by heating the uncondensed gas and therefore making it harder to compress into the core^{3–11}, but such a process seems to require an implausibly large number of supernovae^{6,8,12–14}. Here we show how radiative cooling of intergalactic gas and subsequent supernova heating conspire to eliminate highly compressible low-entropy gas from the intracluster medium. This brings the core entropy and X-ray luminosities of clusters into agreement with the observations, in a way that depends little on the efficiency of supernova heating in the early Universe.

Numerical simulations of cosmological structure formation show that the dark-matter potential wells of clusters are quite similar in shape across a wide range of cluster mass¹⁵. A cluster's mean mass density within the virial radius r_v is roughly 200 times ρ_{cr} , the critical density for closing the Universe, with a dark-matter density distribution $\rho_{DM} \propto [r(1 + cr)^2]^{-1}$, where r is the radius in units of r_v and $c \approx 4–10$ is a parameter governing the concentration of dark matter towards the centre of the cluster. In the absence of radiative cooling and supernova heating, the gas density in such simulated clusters follows nearly the same profile¹⁶, with some additional flattening for $r \lesssim 1/c$.

If all clusters had similar gas density distributions, then cluster temperature would scale as $T \propto \rho_{cr} r_v^2$ and X-ray luminosity would scale as $L \propto \Lambda \rho_{cr}^2 r_v^2$, where Λ is the cooling rate per unit density of the hot gas. At typical cluster temperatures ($T > 2$ keV), free-free emission dominates the cooling, so $\Lambda \propto T^{1/2}$ and we would expect $L \propto T^2$. Instead, the observed power-law index^{17–19} in the relation $L \propto T^b$ is $b \approx 2.6–2.9$, meaning that low-temperature clusters and groups of galaxies are far less luminous than expected. This luminosity deficit is also evident in the low level of the unresolved ~ 1 -keV X-ray background, which is an order of magnitude smaller than predicted by cosmological simulations without radiative cooling or supernova heating^{20–23}.

The X-ray luminosities of low-temperature clusters are unexpectedly small because their gas is less centrally concentrated than in hotter clusters, an effect that has been attributed to a universal minimum entropy level in intracluster gas resulting from supernova heating^{5,9–11}, from heating by active galactic nuclei⁸, or from radiative cooling^{8,12,24,25}. A lower limit to intracluster entropy steepens the relationship between luminosity and temperature, because the shallower potential wells of low-temperature clusters

are less able to overcome the resistance to compression owing to that minimum entropy. Both observations⁵ and theoretical models^{6,10,11,20,21} have established that a core entropy level corresponding to $S \approx 100–200$ keV cm² can account for the observed slope of the $L : T$ relation and the low level of the ~ 1 -keV background.

This entropy scale emerges most naturally from considerations involving radiative cooling. For any cluster of temperature T , we can compute the specific entropy level at which a gas parcel would radiate away its thermal energy in a time equivalent to the age of the Universe. Figure 1 shows that locus in the entropy–temperature plane for a typical cluster heavy-element abundance of 30% of the solar value. The locus for this typical abundance lies in the 100–150 keV cm² range for $T < 2$ keV and rises in proportion to $T^{2/3}$ at higher temperatures. Note that the measured core entropies of clusters and groups³ closely track the cooling locus.

The correspondence between the threshold entropy for cooling (S_c) and observations of core entropy suggests the following picture for the evolution of the intracluster medium. We can consider the medium to consist of independent gas parcels, each with its own entropy history. The luminosity of a sufficiently relaxed cluster, in which entropy increases monotonically with radius, is then completely determined by the shape of its potential well $\rho_{DM}(r)$ and the distribution of specific entropy among its gas parcels. In the context of hierarchical merging, a parcel's entropy history proceeds like this: each merger event affecting the parcel raises its entropy to some new value S_i , and as the gas relaxes after the merger, the parcel's temperature will approach T_i , the characteristic temperature of its new dark-matter halo. A parcel's trajectory in the S – T plane is thus defined by a series of points (S_i, T_i) that proceed upward and to the right in a diagram like Fig. 1.

Now consider the effect of the cooling locus, which creeps

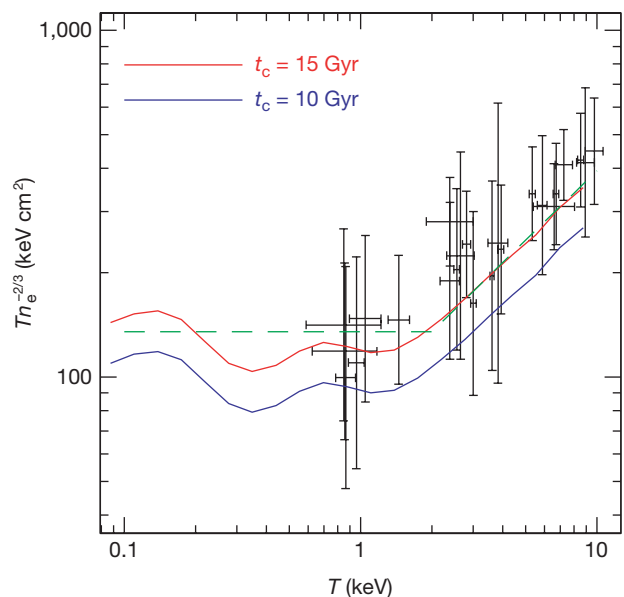


Figure 1 Threshold entropy for cooling within the age of the universe. The solid lines show the specific entropy $S = T n_e^{-2/3}$, where n_e is the electron density, at which gas of temperature T will cool within a time t_c equal to 15 Gyr (red) and 10 Gyr (blue) for elemental abundances of 30% solar. (The usual thermodynamic entropy for an ideal monatomic gas is proportional to $\ln S$.) Note that the core entropy levels in clusters measured at the core radius $0.1 r_v$ (ref. 5) track the cooling threshold for typical cluster heavy-element abundances of 30% solar. Within that core radius, many clusters contain gas whose entropy lies below the cooling threshold, but the luminosity of that relatively small amount of low-entropy gas usually contributes only modestly to the overall luminosity of the cluster¹⁸. The dashed green line indicates the entropy threshold S_c used to determine the $L : T$ relation in Fig. 2.

vertically up the diagram as time progresses. If a parcel's entropy trajectory always remains above this threshold, it will never cool. However, parcels of gas that find themselves below the cooling threshold as structure develops will begin to condense and form stars. Shortly thereafter, supernovae heat the neighbouring gas with an efficiency that remains unknown, adding entropy that tends to suppress further cooling and condensation. Stars will continue to form and supernova will continue to explode until there is no more gas below the cooling threshold; then both processes must cease. Thus cooling and whatever feedback accompanies it act in tandem to eliminate gas parcels with $S < S_c$, inevitably creating an entropy floor at the cooling threshold.

This model leads to an $L : T$ relation that closely matches cluster observations. Suppose that both the gas and dark-matter density are proportional to $[r(1 + cr)^2]^{-1}$ in the absence of cooling. Then we can compute the unmodified entropy distribution $S_g(S)$, the mass of gas with specific entropy less than S , and the inverse relation $S(M_g)$. When cooling is allowed to operate, gas parcels with $S < S_c$ are subject to condensation, and any subsequent feedback is targeted directly at those low-entropy gas parcels. If feedback is inefficient, condensation will remove this gas from the intracluster medium. If feedback is highly efficient, supernova heating will raise the entropy of this gas to $S \gg S_c$, and it will convect to the outer regions of the cluster. In either limit, the modified entropy distribution approaches $S^*(M_g) = S(M_g + M_c)$, where $M_c \equiv M_g(S_c)$. Figure 2 illustrates the $L : T$ relation that we calculate when we allow gas with entropy distribution $S^*(M_g)$ to be in hydrostatic equilibrium with the dark-matter potential, assuming a Hubble constant of

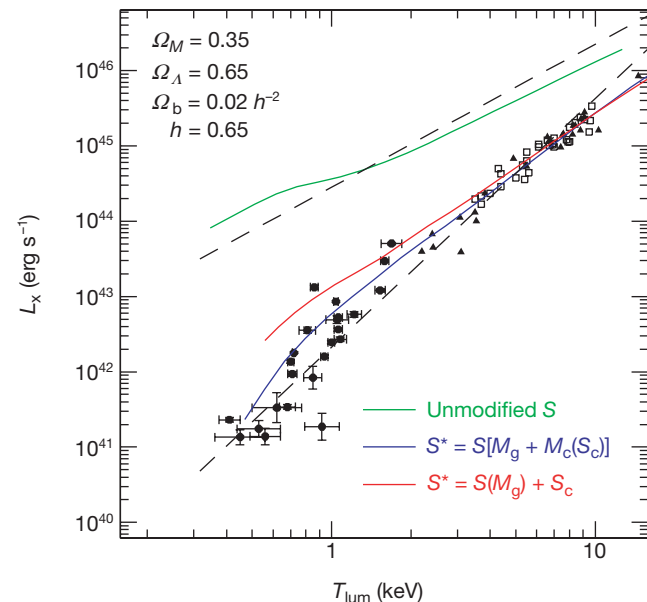


Figure 2 Relation between bolometric X-ray luminosity L_X and luminosity-weighted temperature T_{lum} in clusters and groups of galaxies. Green line, the $L : T$ relation from the unmodified entropy distribution $S(M_g)$ for concentration factors corresponding to model $S_{1,2}$ from ref. 29. Blue line, the $L : T$ relation derived from our primary cooling-threshold model, in which $S^* = S[M_g + M_c(S_c)]$ (see text for details) for the threshold entropy S_c indicated by the dashed green line in Fig. 1. Red line, the $L : T$ relation for an alternative modified-entropy model in which $S^* = S(M_g) + S_c$. Dashed lines, fits to numerical simulations²⁵ of clusters with (lower line) and without (upper line) radiative cooling. Triangles, measurements of L_X and T from a compilation of cluster data¹⁹; open squares, additional cluster data¹⁸; filled circles, group data³⁰. All data and models are normalized to a Hubble constant of $h = 0.65$. The cooling-threshold model naturally accounts for the $L : T$ relation in a standard Λ -dominated cosmology (total matter density, $\Omega_M = 0.35$; dark-energy density, $\Omega_\Lambda = 0.65$; baryon density, $\Omega_b = 0.02h^{-2}$) without any adjustable parameters.

$100 h \text{ km s}^{-1}$ with $h = 0.65$, and a baryon fraction $0.02 h^{-2} \Omega_M^{-1}$ with a total matter density $\Omega_M = 0.33$. (Using an alternative modified entropy distribution, $S^*(M_g) = S(M_g) + S_c$, leads to similar results.)

The $L : T$ relation that we derive from the cooling threshold agrees not only with observations but also with cluster simulations²⁶ that include radiative cooling but no supernova heating. That agreement underlines an important feature of this model: the $L : T$ relation that it predicts is insensitive to the efficiency of supernova heating. Any combination of cooling and feedback that truncates the unmodified entropy distribution at the entropy scale S_c will lead to a similar $L : T$ relation. In that sense, the relation determined by the cooling threshold represents the upper envelope to all $L : T$ relations consistent with the age of the Universe. Extreme amounts of feedback could further lower the luminosity, but apart from relatively small amounts of low-entropy gas that have not yet completely cooled²⁷, the luminosity of a cluster cannot substantially exceed this relation.

Despite the spherical symmetry of the end state, the cooling, condensation and feedback processes that regulate the $L : T$ relation are not restricted to the cluster's core. In hierarchical models of structure formation, the spatial distribution of specific entropy is initially quite complex. If cooling is not allowed to operate, the highest-density, lowest-entropy gas parcels will eventually find their way into the cluster core. However, in the real Universe, this low-entropy gas condenses and forms galaxies long before it reaches the centre of the cluster. Much of the condensation and feedback that ultimately determines a cluster's luminosity therefore pre-dates the epoch of cluster formation. Only modest quantities of gas remain below the cooling threshold in most present-day clusters, where they collect into centrally focused cooling flows.

If this interpretation of the $L : T$ relation is correct, then the level of the entropy floor in clusters ($\sim 100\text{--}200 \text{ keV cm}^2$) conveys little information about the efficiency of supernova heating in the early Universe. That information is more directly reflected in the fraction of baryons that have formed stars. A large proportion of the baryons that are now in clusters were associated with objects of less than 10^{12} solar masses at redshift $z \approx 2$. At that time, the cooling threshold in a dark-matter halo of 10^{12} solar masses would have been $\sim 10 \text{ keV cm}^2$, and virtually all the baryons in those haloes were below that threshold²⁸. Some form of feedback, probably supernovae, prevented most of these baryons from condensing. Once feedback has driven uncondensed gas to high entropy and low density, merger shocks can more easily maintain it above the cooling threshold. Detailed computations, tracing the entropy history of each gas parcel through the series of cooling, feedback and merger events leading to the final cluster, will therefore be needed to relate the supernova heating efficiency to the fraction of baryons now in stars. $\square$

Received 2 July; accepted 26 September 2001.

1. Evrard, A. E. The intracluster gas fraction in X-ray clusters: constraints on the clustered mass density. *Mon. Not. R. Astron. Soc.* **292**, 289–297 (1997).
2. Sarazin, C. L. *X-ray Emissions from Clusters of Galaxies* (Cambridge Univ. Press, Cambridge, 1988).
3. Evrard, A. E. & Henry, J. P. Expectations for X-ray cluster observations by the ROSAT satellite. *Astrophys. J.* **383**, 95–103 (1991).
4. Kaiser, N. Evolution of clusters of galaxies. *Astrophys. J.* **383**, 104–111 (1991).
5. Ponman, T. J., Cannon, D. B. & Navarro, J. F. The thermal imprint of galaxy formation on X-ray clusters. *Nature* **397**, 135–137 (1999).
6. Balogh, M. L., Babul, A. & Patton, D. R. Pre-heated isentropic gas in groups of galaxies. *Mon. Not. R. Astron. Soc.* **307**, 463–479 (1999).
7. Wu, K. K. S., Fabian, A. C. & Nulsen, P. E. J. The effect of supernova heating on cluster properties and constraints on galaxy formation models. *Mon. Not. R. Astron. Soc.* **301**, L20–L24 (1998).
8. Wu, K. K. S., Fabian, A. C. & Nulsen, P. E. J. Non-gravitational heating in the hierarchical formation of X-ray clusters. *Mon. Not. R. Astron. Soc.* **318**, 889–912 (2000).
9. Cavaliere, A., Menci, N. & Tozzi, P. Hot gas in clusters of galaxies: the punctuated equilibria model. *Mon. Not. R. Astron. Soc.* **308**, 599–608 (1999).
10. Tozzi, P. & Norman, C. The evolution of X-ray clusters and the entropy of the intracluster medium. *Astrophys. J.* **546**, 63–84 (2001).
11. Bialek, J. J., Evrard, A. E. & Mohr, J. J. Effects of preheating on X-ray scaling relations in galaxy clusters. *Astrophys. J.* **555**, 597–612 (2001).

12. Bryan, G. L. Explaining the entropy excess in clusters and groups of galaxies without additional heating. *Astrophys. J.* **544**, L1–L5 (2001).
13. Kravtsov, A. V. & Yepes, G. On the supernova heating of the intergalactic medium. *Mon. Not. R. Astron. Soc.* **318**, 227–238 (2000).
14. Bower, R. *et al.* The impact of galaxy formation on the X-ray evolution of clusters. *Mon. Not. R. Astron. Soc.* **325**, 497–508 (2001).
15. Navarro, J. F., Frenk, C. S. & White, S. D. M. A universal density profile from hierarchical clustering. *Astrophys. J.* **490**, 493–508 (1997).
16. Frenk, C. S. *et al.* The Santa Barbara cluster comparison project: a comparison of cosmological hydrodynamics simulations. *Astrophys. J.* **525**, 554–582 (1999).
17. Edge, A. C. & Stewart, G. C. EXOSAT observations of clusters of galaxies. I—The X-ray data. *Mon. Not. R. Astron. Soc.* **252**, 414–427 (1991).
18. Markevitch, M. The L_X -T relation and temperature function for nearby clusters revisited. *Astrophys. J.* **504**, 27–34 (1998).
19. Arnaud, M. & Evrard, A. E. The L_X -T relation and intracluster gas fractions of X-ray clusters. *Mon. Not. R. Astron. Soc.* **305**, 631–640 (1999).
20. Pen, U. Heating of the intergalactic medium. *Astrophys. J.* **510**, L1–L5 (1999).
21. Voit, G. M. & Bryan, G. L. On the distribution of X-ray surface brightness from diffuse gas. *Astrophys. J.* **551**, L139–L142 (2001).
22. Bryan, G. L. & Voit, G. M. The X-ray surface brightness distribution from diffuse gas. *Astrophys. J.* **556**, 590–600 (2001).
23. Wu, K. K. S., Fabian, A. C. & Nulsen, P. E. J. The soft X-ray background: evidence for widespread disruption of the gas halos of galaxy groups. *Mon. Not. R. Astron. Soc.* **324**, 95–107 (2001).
24. Knight, P. A. & Ponman, T. J. The properties of the hot gas in galaxy groups and clusters from 1D hydrodynamical simulations—I. Cosmological infall models. *Mon. Not. R. Astron. Soc.* **289**, 955–972 (1997).
25. Pearce, F. R., Thomas, P. A., Couchman, H. M. P. & Edge, A. C. The effect of radiative cooling on the X-ray properties of galaxy clusters. *Mon. Not. R. Astron. Soc.* **317**, 1029–1049 (2000).
26. Muanwong, O., Thomas, P. A., Kay, S. T., Pearce, F. R. & Couchman, H. M. P. The effect of radiative cooling on scaling laws of clusters and groups. *Astrophys. J.* **552**, L27–L30 (2001).
27. Fabian, A. C., Crawford, C. S., Edge, A. C. & Mushotsky, R. F. Cooling flows and the X-ray luminosity-temperature relation for clusters. *Mon. Not. R. Astron. Soc.* **267**, 779–784 (1994).
28. Balogh, M., Pearce, F. R., Bower, R. G. & Kay, S. T. Revisiting the cosmic cooling crisis. *Mon. Not. R. Astron. Soc.* **326**, 1228–1234 (2001).
29. Eke, V., Navarro, J. F. & Steinmetz, M. The power spectrum dependence of dark matter halo concentrations. *Astrophys. J.* **554**, 114–125 (2001).
30. Helsdon, S. F. & Ponman, T. J. The intragroup medium in loose groups of galaxies. *Mon. Not. R. Astron. Soc.* **315**, 356–370 (2000).

Acknowledgements

We thank R. Bower and M. Balogh for discussions. G.M.V. receives partial support from NASA, and G.L.B. is supported by NASA through a Hubble fellowship.

Correspondence and requests for materials should be addressed to G.M.V. (e-mail: voit@stsci.edu).

Non-Fermi-liquid nature of the normal state of itinerant-electron ferromagnets

C. Pfleiderer*, S. R. Julian† & G. G. Lonzarich†

* *Physikalisches Institut, Universität Karlsruhe, D-76128 Karlsruhe, Germany*

† *Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK*

A century of research on magnetic phenomena had led to the view that the normal state of itinerant-electron ferromagnets such as Fe, Ni and Co could be described in terms of the standard model of the metallic state or its extension known as the nearly ferromagnetic Fermi liquid theory^{1–3}. In recent years, however, a large body of observations has accumulated from various complex intermetallic systems^{4,5} that raises the possibility that this assumption might be wrong. Here we examine this issue by means of high-precision measurements of the electrical transport and magnetic properties of pure ferromagnets—in particular, MnSi—in which the Curie temperature is tuned towards absolute zero by the application of hydrostatic pressure. With this method, it is possible for us to study the normal state over an extraordinarily large range of temperature of up to five orders of magnitude above the Curie temperature. Our results using MnSi reveal a

particularly striking combination of properties—most notably a $T^{3/2}$ power law for the resistivity—showing clearly that the normal state of this itinerant-electron ferromagnet cannot be described in terms of the standard model of metals.

As in the related investigations of the puzzling normal-state properties of the high-temperature superconductors, the clarification of the normal state of elemental itinerant-electron ferromagnets is hampered by the relatively high ordering temperatures, because a range of complicating effects can arise. Further difficulties are in many cases related to subtle complexities of the electronic state in the presence of uncontrolled forms of structural and chemical disorder and of crystalline and band structure anisotropies that can lead to ambiguities in the effective dimensionality of the systems. To minimize these complications we have restricted ourselves to stoichiometric ferromagnets that can be produced in a pure state and in which the Curie temperature is low and can be suppressed to absolute zero by the application of hydrostatic pressure of moderate intensity. Moreover, we have selected relatively uncomplicated cubic systems that are unambiguously three-dimensional in which the electronic and magnetic properties are essentially isotropic.

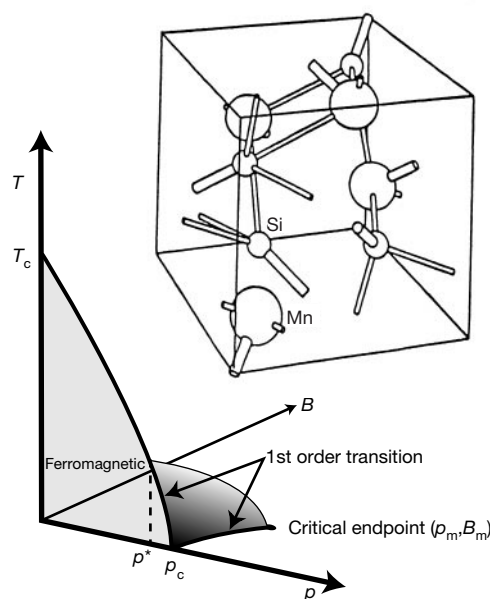


Figure 1 Diagram of the ferromagnetic state of MnSi and its crystallographic structure^{16,17}. In the temperature–pressure (T – p) plane ($B = 0$) the transition drops with increasing pressure from its ambient pressure value of $T_c = 29.5$ K and changes from second to weakly first order at $p^* = 12$ kbar where $T_c \approx 12$ K before it disappears above $p_c = 14.6$ kbar. In the magnetic field–pressure (B – p) plane ($T = 0$) the transition is weakly first-order up to a critical endpoint, estimated to be at $B_m = 0.6$ T and $p_m = 17$ kbar. In the dome to the left MnSi is ordered ferromagnetically. The inset shows the crystallographic unit cell of the B20 crystal structure of MnSi. We note the absence of inversion symmetry, which we take to explain the absence of superconductivity that would normally be expected at the border of itinerant ferromagnetism^{11,12}. This and the well-developed first-order regime near p_c make the deviations from Fermi-liquid behaviour observed in MnSi particularly striking for the high sample purity investigated, because they may be tracked down to the low-millikelvin range. The single crystal samples investigated here were grown by radio-frequency induction melting of zone-refined high-purity starting materials on a water-cooled Cu crucible in an ultrahigh vacuum environment. The resulting ingots were annealed for two weeks near the melting point. Single crystal samples were oriented by Laue X-ray diffraction and cut by low-power spark erosion. The sample investigated here exhibits quantum oscillations at ambient pressure consistent with charge carrier mean free paths around $5,000$ Å in agreement with $\rho_0 = 0.33$ $\mu\Omega$ cm (ref. 20). Further evidence of the high sample quality include, for instance, a very low mosaic spread observed in X-ray diffraction, and confirmation of the well documented ambient pressure magnetic properties by elastic neutron scattering.

Here we have studied a system in which the deviations from the predictions of the standard model are particularly striking: MnSi. The ordering temperature of MnSi is $T_c = 29.5$ K at ambient pressure and vanishes above a critical pressure, p_c , of only 14.6 kbar. MnSi is perhaps the most extensively studied of the itinerant-electron ferromagnets, apart from the elemental metals

iron, cobalt and nickel, which have high T_c and correspondingly very high critical pressures. It is a stable, congruently melting compound that can be produced in ultra-pure form and crystallizes in the cubic B20 structure shown in the inset of Fig. 1 (ref. 6). Although it is relatively simple, the B20 structure is somewhat unusual in that it lacks strict space-inversion symmetry. This leads to a well understood long-wavelength helical twist in the ferromagnetic polarization (refs 7–10). More importantly, for the purposes of this study, the broken inversion symmetry may also be instrumental in suppressing superconductivity that would otherwise be expected to arise on the border of itinerant-electron ferromagnetism^{11,12}. This gives us the opportunity to track the normal metallic state of interest near p_c over five orders of magnitude in temperature, from below 10 mK to room temperature, in an ultra-pure system without intervening phase transitions.

The large body of information available on MnSi includes thermodynamic and transport data as a function of temperature, magnetic field and pressure^{13–17}, and microscopic data based on neutron scattering¹⁸, nuclear magnetic resonance¹⁹ and quantum oscillatory phenomena²⁰. The experimental information and the fact that MnSi satisfies the requirements of simplicity described above, as well as numerous relevant theoretical investigations^{5,11,21,24},

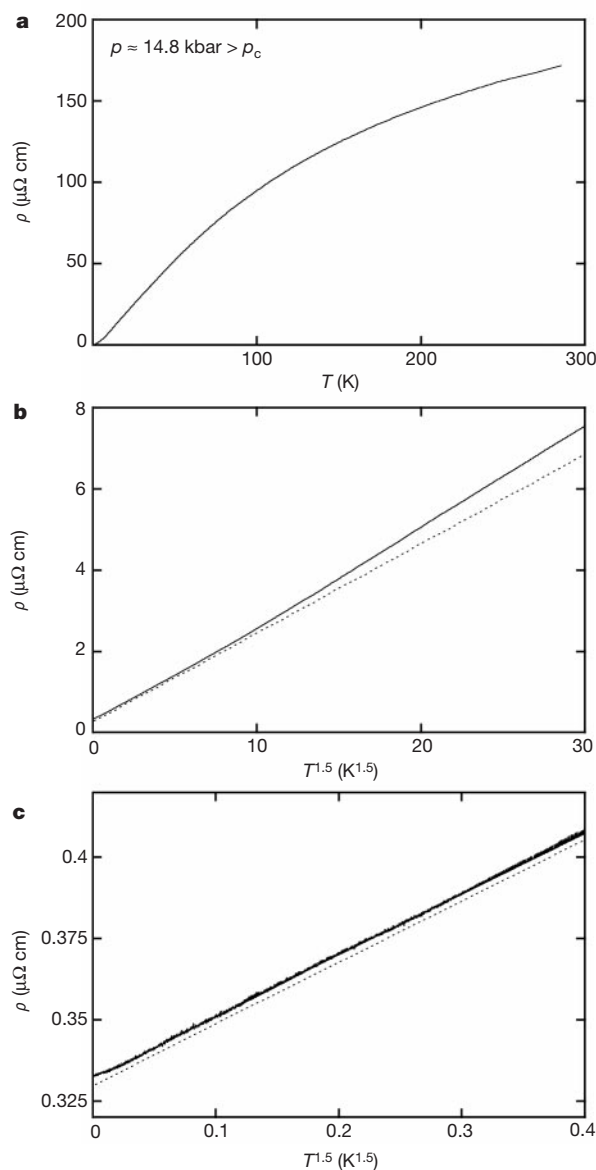


Figure 2 Temperature dependence of the resistivity $\rho(T)$ of MnSi above the critical pressure in different T regimes. **a**, The resistivity on a linear T scale below room temperature. $\rho(T)$ drops monotonically from a high T value by nearly three orders of magnitude and changes from a sublinear to a superlinear dependence below roughly 15 K. **b**, The resistivity on a $T^{3/2}$ scale below 10 K. Below 5 K a $T^{3/2}$ dependence is observed, where the T -dependent part is large compared to ρ_0 . Dotted line represents a $T^{3/2}$ behaviour. **c**, The resistivity on a $T^{3/2}$ scale below 0.5 K. A $T^{3/2}$ dependence is observed down to the lowest T investigated, where the T -dependent part is small compared to ρ_0 . Dotted line represents a $T^{3/2}$ behaviour. For the high pressure experiments, a non-magnetic miniature clamp cell was developed that is reduced in weight by a factor of 100 compared with conventional apparatus to ensure optimal cooling. The pressure cell was attached to a Cu nuclear demagnetization stage that was pre-cooled by a dilution refrigerator. Cryogenically cooled detection electronics allowed measurements at very low excitation levels even in the low-millikelvin range. Thermal equilibrium between sample and thermometers was confirmed by a thermocouple measured with a superconducting quantum interference device (SQUID) to the lowest temperatures.

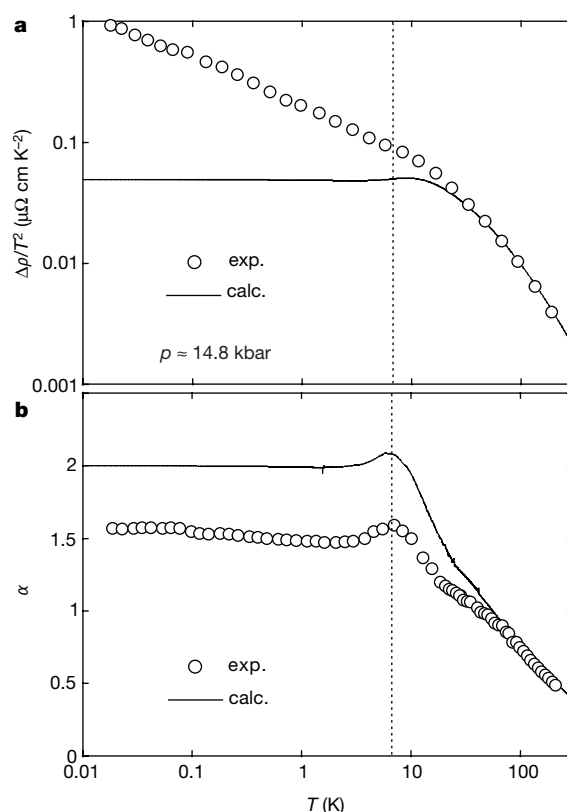


Figure 3 Comparison of the experimentally observed temperature dependence of the resistivity above $p_c = 14.6$ kbar with predictions of the model of a nearly ferromagnetic Fermi liquid (NFFL)^{5,11,16,21,24,25} as evaluated for MnSi. **a**, Temperature-dependent part of the resistivity divided by the square of the temperature $\Delta\rho(T)/T^2$. In Fermi-liquid theory this quantity measures the quasiparticle transport scattering cross-section. Experimentally $\Delta\rho(T)/T^2$ is found to diverge down to the lowest temperatures. In the NFFL model a Fermi-liquid form is expected already as high as a few kelvin. **b**, The power-law exponent α , extracted from $\Delta\rho(T) \propto T^\alpha$, versus temperature on a logarithmic scale. The experimentally observed exponent crosses from the high- T to the low- T regime around 15 K. Below several kelvin it locks to a $T^{3/2}$ dependence over nearly three orders of magnitude in T . The expected behaviour in the NFFL model as evaluated for MnSi predicts a T^2 dependence below a few kelvin. We note that the crossover towards the high- T behaviour is predicted correctly in contrast with the qualitative $T^{3/2}$ form at low T , which is radically different.

provide us with a solid foundation from which we re-examine closely the applicability of Fermi-liquid theory for the normal state above T_c for a pure itinerant-electron ferromagnet in which T_c can be tuned to absolute zero.

In the phase diagram of temperature versus pressure and field, the Curie temperature, T_c , in MnSi falls monotonically with pressure, p , and magnetic field, B , as shown schematically in Fig. 1 (refs 16, 17). The transition is second-order along the zero-field axis up to $p^* \approx 12$ kbar where $T_c \approx 12$ K and along the contour at finite fields terminating at the point ($p_m \approx 17$ kbar, $B_m \approx 0.6$ T) where T_c vanishes. The transition is, however, weakly first-order within a cavity extending to finite fields, and in particular between p^* and p_c where T_c falls to zero along the zero-field axis. For a fixed field, a first-order line in the B - p plane is seen to end in a second-order point in a way which is reminiscent of a conventional liquid-gas transition in the van der Waals model. At B_m this line collapses to a single point that represents a second-order transition with zero T_c . This type of phase diagram has now been observed in several related materials^{22,23} and theoretical considerations suggest that it may be generally valid for itinerant-electron ferromagnets in the limit $T_c \rightarrow 0$.

We shall focus our attention on the temperature dependence of the electrical resistivity in the p - B region of the phase diagram just outside the dome in which magnetic order is found. Our main result is that the resistivity in this region exhibits a temperature dependence of the form $\rho(T) = \rho_0 + AT^{3/2}$ over a remarkably wide range in temperature of nearly three orders of magnitude. An example of this striking behaviour is shown in Figs 2 and 3 for a pressure just above p_c in the Earth's magnetic field and for a pure sample with a residual resistivity ρ_0 as low as $0.33 \mu\Omega \text{ cm}$. As illustrated in Fig. 4, this behaviour extends also up to the boundary of ferromagnetism as a function of field for pressures close to but above p_c .

The $T^{3/2}$ power law for the resistivity is at odds with the T^2 temperature dependence expected in Fermi-liquid theory (Fig. 3). This conventional T^2 form of the resistivity is indeed observed in MnSi but only at low temperatures in (1) the ferromagnetic state (inside the dome of the phase diagram) and (2) the normal state at pressures far above p_c . We find that the $T^{3/2}$ coefficient, A , is only weakly dependent on ρ_0 in the range 0.35 – $4.0 \mu\Omega \text{ cm}$, but is sensitive to pressure and falls by approximately a factor of 2 from p_c to $2p_c$. We have also observed $T^{3/2}$ variations of the resistivity in two other itinerant-electron ferromagnets with cubic lattices, ZrZn_2 (C.P., M. Uhlarz, H. v. Löhneysen and S. M. Hayden, unpublished work) and Ni_3Al (I. R. Walker, N. R. Bernhoeft and

G.G.L., unpublished work). This suggests that the $T^{3/2}$ power law may have a general origin for the class of systems we are considering and is not a consequence of some peculiarity of MnSi alone.

In the search for a possible explanation of the $T^{3/2}$ power law we consider first the predictions of a theoretically well established extension of the Fermi-liquid picture to which we shall refer as the nearly ferromagnetic Fermi-liquid (NFFL) model^{5,11,16,21,24,25}. The behaviour of the NFFL model depends on the temperature scales $T^* = T_F/(k_F\xi S)$, where T_F is the Fermi temperature, k_F is the Fermi wavevector, ξ is the magnetic correlation length and S is the Stoner enhancement factor (the ratio of the magnetic susceptibility with and without exchange enhancement). For temperatures well below T^* , the NFFL model reduces to the Fermi-liquid model which is characterized by a T^2 resistivity, as already noted. Above T^* , on the other hand, the NFFL model reduces to the so-called marginal Fermi-liquid model^{11,16,21,24} which is characterized by a $T^{5/3}$ resistivity when T is far below T_F for a three-dimensional system and for a dominance of small-angle electron scattering in the electronic transport. We note that when the latter assumption breaks down $\rho(T)$ is expected to be linear in T . This is also the form predicted by a different kind of marginal Fermi-liquid theory introduced to account for the normal-state behaviour of the high-temperature superconductors²⁶.

A comparison of the prediction of the NFFL model with our experiment is shown in Fig. 3. At p_c and zero magnetic field, T^* for MnSi is consistent within a few K with the first-order nature of T_c just below p_c . The resistivity exponent is thus expected to be described by the marginal Fermi-liquid model above 5 K, but by Fermi-liquid theory in the millikelvin range. As shown in Fig. 3, the NFFL model does indeed agree with experiment above a few kelvin but is seen to break down dramatically at lower temperatures where the unexpected $T^{3/2}$ power law prevails.

We note that the $T^{3/2}$ power law qualitatively corresponds to the effects of frozen-in disorder, as observed, for example, in amorphous metals and polycrystalline spin glasses²⁷, which are expected to lead to a diffusive motion of charge carriers on length scales substantially larger than the mean free path l (refs 28, 29). But the single-crystal samples investigated here exhibit, at ambient pressure, quantum oscillations corresponding to l of the order of $5,000 \text{ \AA}$, in agreement with ρ_0 (ref. 20). Thus, only well below a new temperature scale $T_F/(k_F l)$ it is possible that a $T^{3/2}$ term in the resistivity may arise. This scale, however, is at least an order of magnitude too small to explain the range of the $T^{3/2}$ resistivity shown in Figs 2 and 3. Apart from this, the impurity model would also seem to be at odds with the observed weak dependence of A on ρ_0 and the fact that the

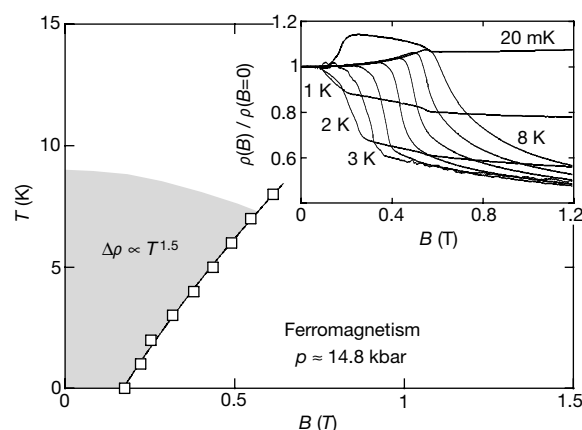


Figure 4 Regime of $T^{3/2}$ resistivity in the T - B plane just above $p_c = 14.6$ kbar. The normalized longitudinal magnetoresistance $\rho(B)/\rho(B=0)$, shown in the inset, scales below the transition to the ferromagnetic state, marked by a pronounced drop. This implies that the T dependence of the resistivity on the low-field side keeps an unchanged

$T^{3/2}$ dependence as shown in the main figure, of which the coefficient A is only very weakly field-dependent. Data in the inset between 3 K and 8 K correspond to temperatures of 4 K, 5 K, 6 K and 7 K, respectively.

first-order transition does not exhibit any broadening as a function of pressure¹⁶.

We conclude that the $T^{3/2}$ power law of the resistivity in MnSi and in related pure systems (under investigation) is not consistent with current models of itinerant-electron ferromagnetism. The origin of this form of $\rho(T)$ may lie in extremely subtle effects of disorder in the highest-purity samples, pointing at an entirely unexpected and new aspect of the metallic state. It is more likely, however, to lie in a diffusive motion of the electrons resulting from the interactions among the itinerant electrons themselves. This constitutes a breakdown of the Fermi-liquid or NFFL model, long assumed to be valid for the normal state of itinerant-electron ferromagnets. □

Received 2 August; accepted 12 September 2001.

1. Stoner, E. C. Collective electron ferromagnetism. *Proc. R. Soc. Lond. A* **165**, 372–414 (1938).
2. Radu, G. T. & Suhl, H. (eds) *Magnetism—A Treatise on Modern Theory and Materials* (Academic, New York, 1966).
3. Lonzarich, G. G. in *Electrons at the Fermi Surface* (ed. Springford, M.) 225–528 (Cambridge Univ. Press, Cambridge, 1980).
4. *Institute of Physics Conference on Non-Fermi Liquid Behaviour in Metals*. *J. Phys. Cond. Matter*. (Spec. Issue) **8**, 9675–10148 (1996).
5. Varma, C. M., Nussinov, Z. & van Saarloos, W. Singular Fermi liquids. *Phys. Rep.* (in the press); also preprint cond-mat 0103393 at (<http://xxx.lanl.gov>).
6. Williams, H. J., Wernick, J. H., Sherwood, R. C. & Wertheim, G. K. Magnetic properties of monosilicides of some 3-d transition elements. *J. Appl. Phys.* **37**, 1256 (1966).
7. Plumer, M. L. & Walker, M. B. Wavevector and spin reorientation in MnSi. *J. Phys. C: Solid State Phys.* **14**, 4689–4699 (1981).
8. Ishikawa, Y. & Arai, M. Magnetic phase diagram of MnSi near critical temperature studied by neutron small angle scattering. *J. Phys. Soc. Jpn* **53**, 2726–2733 (1984).
9. Ishida, M. *et al.* Crystal chirality and helicity of the helical spin density wave in MnSi: II. polarized neutron diffraction. *J. Phys. Soc. Jpn* **54**, 2975–2982 (1985).
10. Lebech, B. in *Recent Advances in Magnetism of Transition Metal Compounds* 167–178 (World Scientific, Singapore, 1993).
11. Lonzarich, G. G. in *Electron* (ed. Springford, M.) 109–147 (Cambridge Univ. Press, Cambridge, 1997).
12. Pfeiderer, C. *et al.* Coexistence of superconductivity and ferromagnetism in the d-band metal $ZrZn_2$. *Nature* **412**, 58–61 (2001); Erratum *Nature* **412**, 660 (2001).
13. Fawcett, E., Maita, J. P. & Wernick, J. H. Magnetoelastic and thermal properties of MnSi. *Int. J. Magn.* **1**, 29–34 (1970).
14. Bloch, D., Voiron, J., Jaccarino, V. & Wernick, J. H. The high field–high pressure magnetic properties of MnSi. *Phys. Lett. A* **51**, 259–291 (1975).
15. Thompson, J. D., Fisk, Z. & Lonzarich, G. G. Perspective on heavy-electron and Kondo-lattice systems from high pressure studies. *Physica B* **161**, 317–323 (1989).
16. Pfeiderer, C., McMullan, G. J., Julian, S. R. & Lonzarich, G. G. Magnetic quantum phase transition in MnSi under hydrostatic pressure. *Phys. Rev. B* **55**, 8330–8338 (1997).
17. Thessieu, C. *et al.* Field dependence of the magnetic quantum phase transition in MnSi. *J. Phys. Condens. Matter* **9**, 6677–6687 (1997).
18. Ishikawa, Y. *et al.* Paramagnetic spin fluctuations in the weak itinerant-electron ferromagnet MnSi. *Phys. Rev. B* **31**, 5884–5893 (1985).
19. Yasuoka, H., Jaccarino, V., Sherwood, R. C. & Wernick, J. H. NMR and susceptibility studies of MnSi above T_c . *J. Phys. Soc. Jpn* **44**, 842–849 (1978).
20. Taillefer, L., Lonzarich, G. G. & Strange, P. The band magnetism of MnSi. *J. Magn. Magn. Mater.* **54–57**, 957–958 (1986).
21. Millis, A. J. Effect of a nonzero temperature on quantum critical points in itinerant fermion systems. *Phys. Rev. B* **48**, 7183–7196 (1993).
22. Goto, T., Shindo, Y., Takahashi, H. & Ogawa, S. Magnetic properties of the itinerant metamagnetic system $\text{Co}(\text{S}_{1-x}\text{Se}_x)_2$ under high magnetic field and high pressure. *Phys. Rev. B* **56**(21), 14019–14028 (1997).
23. Huxley, A., Sheikin, I. & Braithwaite, D. Metamagnetic behavior near the quantum critical point in UGe_2 . *Physica B* **284–288**, 1277–1278 (2000).
24. Moriya, T. *Spin Fluctuations in Itinerant Electron Magnetism* (Springer, Berlin, 1985).
25. Lonzarich, G. G. & Taillefer, L. Effect of spin fluctuations on the magnetic equation of state of ferromagnetic or nearly ferromagnetic metals. *J. Phys. C* **18**, 4339–4371 (1985).
26. Varma, C. M. *et al.* Phenomenology of the normal state of Cu–O high-temperature superconductors. *Phys. Rev. Lett.* **63**, 1996–1999 (1989).
27. Ford, P. & Mydosh, J. A. Electrical resistivity of noble-metal-host-3d solute spin glass alloys. *Phys. Rev. B* **14**, 2057–2070 (1976).
28. Rivier, N. & Mensah, A. E. Low temperature resistivity and collective excitations. *Physica B* **91**, 85–88 (1977).
29. Fischer, K. H. On the electrical resistivity of spin glasses. *Z. Phys. B* **34**, 45–53 (1979).

Acknowledgements

We would like to thank N. R. Bernhoeft, S. Brown, P. Coleman, N. Doiron-Layraud, J. Flouquet, F. M. Grosche, S. M. Hayden, D. Khmel'nitskii, H. v. Löhneysen, G. J. McMullen, A. J. Millis, A. Rosch, S. Sachdev, L. Taillefer, A. Tsvetlik, T. Vojta and I. R. Walker. Financial support by the Deutsche Forschungsgemeinschaft (Germany), the Engineering and Physical Sciences Research Council (UK) and the European Science Foundation under FERLIN are gratefully acknowledged.

Correspondence and requests for materials should be addressed to C.P. (e-mail: christian.pfeiderer@physik.uni-karlsruhe.de).

Programmable and autonomous computing machine made of biomolecules

Yaakov Benenson^{*†}, Tamar Paz-Elizur[†], Rivka Adar[†], Ehud Keinan^{‡§}, Zvi Livneh[†] & Ehud Shapiro^{*†}

^{*} Department of Computer Science and Applied Mathematics, [†] Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel
[‡] Department of Chemistry and Institute of Catalysis Science and Technology Technion – Israel Institute of Technology, Haifa 32000, Israel
[§] Department of Molecular Biology and Skaggs Institute of Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA

Devices that convert information from one form into another according to a definite procedure are known as automata. One such hypothetical device is the universal Turing machine¹, which stimulated work leading to the development of modern computers. The Turing machine and its special cases², including finite automata³, operate by scanning a data tape, whose striking analogy to information-encoding biopolymers inspired several designs for molecular DNA computers^{4–8}. Laboratory-scale computing using DNA and human-assisted protocols has been demonstrated^{9–15}, but the realization of computing devices operating autonomously on the molecular scale remains rare^{16–20}. Here we describe a programmable finite automaton comprising DNA and DNA-manipulating enzymes that solves computational problems autonomously. The automaton's hardware consists of a restriction nuclease and ligase, the software and input are encoded by double-stranded DNA, and programming amounts to choosing appropriate software molecules. Upon mixing solutions containing these components, the automaton processes the input molecule via a cascade of restriction, hybridization and ligation cycles, producing a detectable output molecule that encodes the automaton's final state, and thus the computational result. In our implementation 10^{12} automata sharing the same software run independently and in parallel on inputs (which could, in principle, be distinct) in 120 μl solution at room temperature at a combined rate of 10^9 transitions per second with a transition fidelity greater than 99.8%, consuming less than 10^{-10} W.

A finite automaton³ is a notional computing machine that operates on finite sequences of symbols. The machine can be in one of a finite number of internal states, of which one is designated an initial state and some are designated accepting states. Its software consists of transition rules, each specifying a next state based on the current state and the current symbol. It is initially positioned on the leftmost input symbol in the initial state. In each transition the machine moves one symbol to the right, changing its internal state according to one of the applicable transition rules. Alternatively, it may 'suspend' without completing the computation if no transition rule applies. A computation terminates on processing the last input symbol. An automaton is said to accept an input if a computation on this input terminates in an accepting final state.

The basic features and processes of a finite automaton with two internal states (S0 and S1) and an alphabet comprising two input symbols a and b are shown in Fig. 1. This automaton can have eight possible transition rules (T1–T8; Fig. 1c), and programming amounts to selecting some of these transition rules and deciding which internal states are accepting. There are 255 possible transition-rule selections and 3 possible selections of accepting states (either S0 or S1, or both), resulting in 765 syntactically distinct programs. A selection of such programs (A2–A7) is shown in Fig. 1d.

The design of our molecular finite automaton incorporates ideas

from designs for molecular Turing machines^{5,8}. Its hardware consists of a mixture of the class IIS restriction nuclease *FokI*, T4 DNA ligase and ATP, while the software comprises eight short double-stranded (ds) DNA molecules, the 'transition molecules' (Fig. 2a), which encode the eight possible transition rules (Fig. 1c). A dsDNA molecule encodes the initial state of the automaton and the input (Fig. 2b), with six base pairs (bp) coding for one input symbol (Fig. 2c). The system also contains 'peripherals', two output-detection molecules of different lengths (Fig. 2d), each of which can interact selectively with a different output molecule to form an output-reporting molecule that indicates a final state and can be readily detected by gel electrophoresis. The computation starts when the hardware, software and input are all mixed together and runs autonomously, if possible till termination. If the peripherals are also mixed then output reporters are formed *in situ* on termination.

The automaton processes the input as shown in Fig. 2e. First, the input is cleaved by *FokI*, thereby exposing a four-nucleotide sticky end that encodes the initial state and the first input symbol. The computation proceeds via a cascade of transition cycles. In each cycle the sticky end of an applicable transition molecule ligates to the sticky end of the input molecule, detecting the current state and the current symbol. The product is cleaved by *FokI* inside the next symbol encoding, exposing a new four-nucleotide sticky end. The design of the transition molecules ensures that the 6-bp-long encodings of the input symbols a and b are cleaved by *FokI* at only two different 'frames'⁵, the leftmost frame encoding the state S1 and the rightmost frame encoding S0 (Fig. 2c). The exact next restriction site and thus the next internal state are determined by the current state and the size of the spacers (Fig. 2a, green) in an applicable transition molecule. The computation proceeds until no transition molecule matches the exposed sticky end of the input or until the special terminator symbol is cleaved, forming an output molecule that has a sticky end encoding the final state. In a step extraneous to the computation and analogous to a 'print' instruction of a conventional computer, this sticky end ligates to one of two output detectors and the resultant output reporter is identified by gel electrophoresis.

We tested the operation of the automaton by running it with a selection of programs on various inputs (Fig. 3a). Each lane in the gel-electrophoresis image presents the result of running an indicated program on a specified input, showing the predicted output reporters in all cases or suspending where expected. The incorrect formation of a small amount of S1-R in an A6-aba computation and the formation of an unidentified product in an A5-aba computation are under investigation. Figure 3b shows computations of automaton A1 (Fig. 1a) on 6-symbol-long inputs, producing the expected outputs.

In all reactions, each input molecule initiates an independent parallel computation. Thus, the results depicted in Figs 3a and b represent 10¹² automata operating in parallel. To illustrate this independence, we ran the same program on a mixture of two different inputs. Figure 3c shows that both output reporters were formed when expected. In the case of expected suspension on one of the inputs, a single correct reporter was formed.

We also tested a non-deterministic automaton A7 (Fig. 1d). Only computations on input aabb reached the accepting state S0, though some computations on this input reached state S1 and some suspended owing to different non-deterministic choices, as expected (Fig. 3d). Although the deterministic equivalent of A7 requires three states and cannot be programmed in our system, using a non-deterministic approach in real applications is disadvantageous as the yield decreases exponentially with the number of non-deterministic choices made during a computation.

We validated the mechanism of operation (Fig. 2e) by the results of three experiments: omission of single components, analysis of the intermediates, and measurement of computation fidelity (Fig. 4).

First, we ran A1 on ababaa, omitting a single reaction component at a time (Fig. 4a). No output reporters could be detected once an essential component was omitted except for the case of T6 removal, an incorrectness similar to the case of the A6-aba computation. In addition, omitting ligase did not seem to impair the formation of the output molecule, although without ligase the reporter could not be detected. The possibility of ligase-free computing is under investigation.

In the second experiment, we analysed the sizes of computation

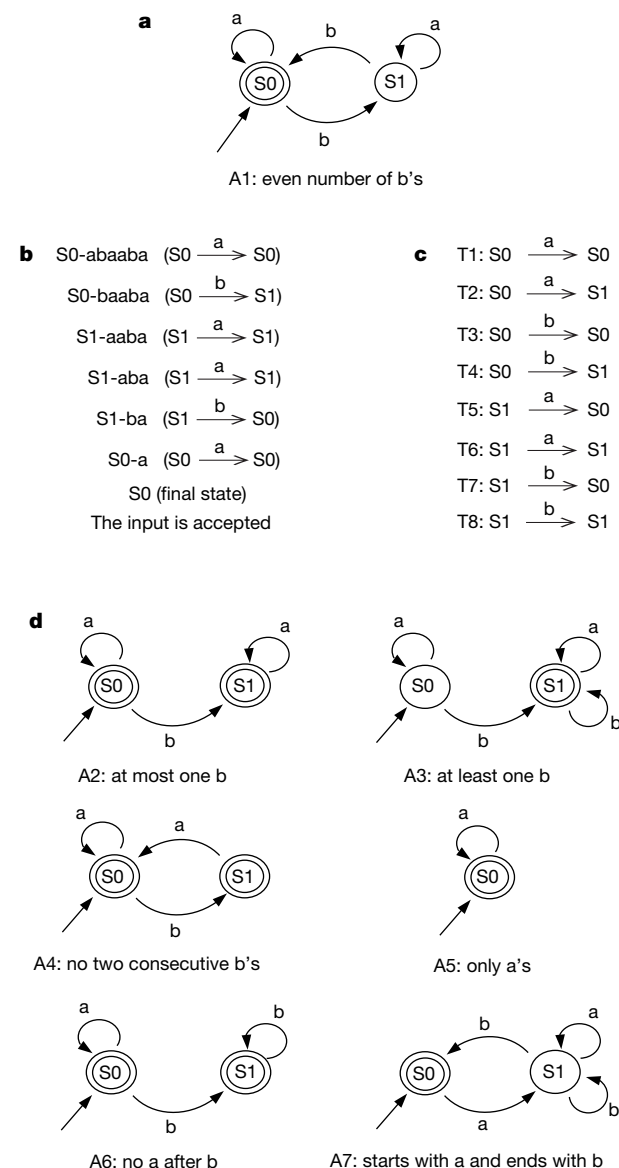


Figure 1 Finite automata with two states (S0 and S1) and two symbols (a and b). **a**, Diagram representing the automaton A1 accepting inputs with an even number of b symbols. Incoming unlabelled arrow represents the initial state, labelled arrows represent transition rules, and the double circle represents an accepting state. **b**, An example computation over abaaba. Each row represents an intermediate configuration, showing the current state of the automaton and the remaining symbols to be processed. The transition rule taking a configuration to its successor is shown on the right. **c**, A list of all eight possible transition rules of a two-state two-symbol automaton. **d**, Six other automata programs used to test the molecular implementation and the sets of inputs they accept. Non-deterministic automaton A7 has two transitions T7 and T8 applicable to the same configuration. A computation of A7 ending in an accepting state uses T8 for all b symbols except the last, and uses T7 for the last b.

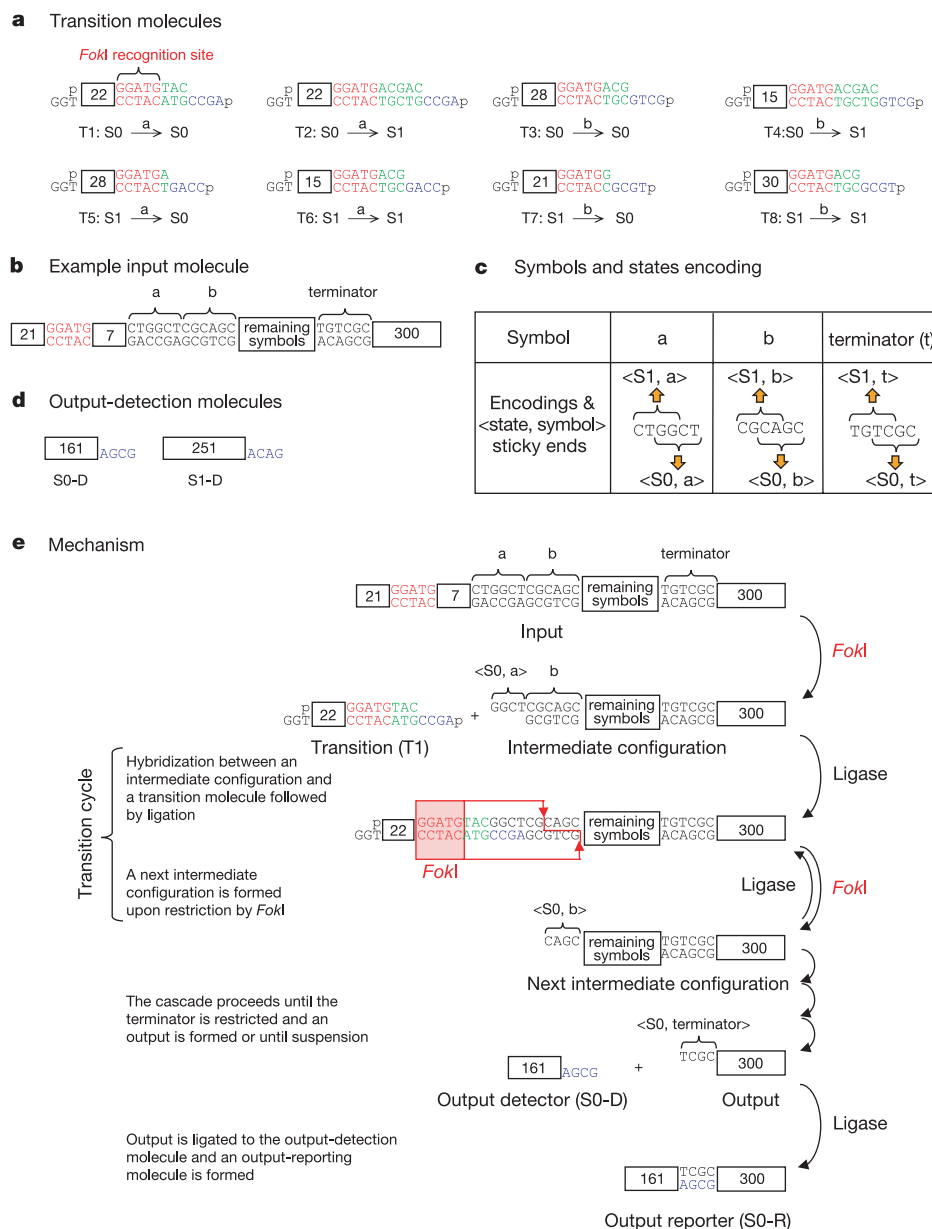


Figure 2 Design details and mechanism of operation of the molecular finite automaton. **a**, Structure of the transition molecules. A transition molecule detects the current state and symbol and determines the next state. It consists of {state, symbol} detector (blue), *FokI* recognition site (red) and spacer (green) that determines the location of the *FokI* cleavage site inside the next symbol encoding, which in turn defines a next state. 1-bp spacers effect S1 to S0 transition, 3-bp maintain the current state, and 5-bp transfer S0 to

S1. **b**, Structure of an input molecule. **c**, The encoding for the input symbols a, b, and terminator (sense strands) and the sequences of the {state, symbol} sticky ends. **d**, Structure of the output-detection molecules. **e**, A sample processing of an input molecule beginning with ab. The S0 $\xrightarrow{a}$ S0 rule is applied to the first intermediate configuration molecule. The detection is illustrated for S0 output.

intermediates (Fig. 4b). The A1 program with input ababaa uses all its transition molecules in the order T1, T4, T6, T7 in the first four cycles. Thus, removal of each transition molecule would suspend the computation in each of its first four intermediates. These molecules would accumulate, 'waiting' for hybridization with the missing transition molecule, together with the product of the reverse ligation with a fragment restricted at the previous cycle (Fig. 2e). This experiment was performed with a ^{32}P -labelled ababaa input. The expected sizes of the intermediates were compared to the actual sizes obtained using a calibration curve based on the standards ladder. The correctness of this measurement is $\pm 1\%$ (± 3.5 bp). The sizes of the 'restricted' (lower) bands are shown in the following (measured, predicted) pairs for each removed

transition molecule: (-T1): (365 ± 4 , 360); (-T4): (357 ± 4 , 354); (-T6): (352 ± 4 , 350); (-T7): (344 ± 4 , 344). The differences between the lengths of the intermediates are 8, 5 and 8 bp between the first and second, second and third, and third and fourth intermediates, respectively. The predicted values are 6, 4 and 6 bp, respectively. These data support the proposed mechanism within the experimental error. More precise experiments are under way.

In the third experiment, we measured the computation fidelity by increasing the detection sensitivity using ^{32}P -labelled output detectors. In two computations with opposite predicted outputs, no incorrect products could be detected even after prolonged exposure. Given the detection limit, the overall fidelity of the computations is

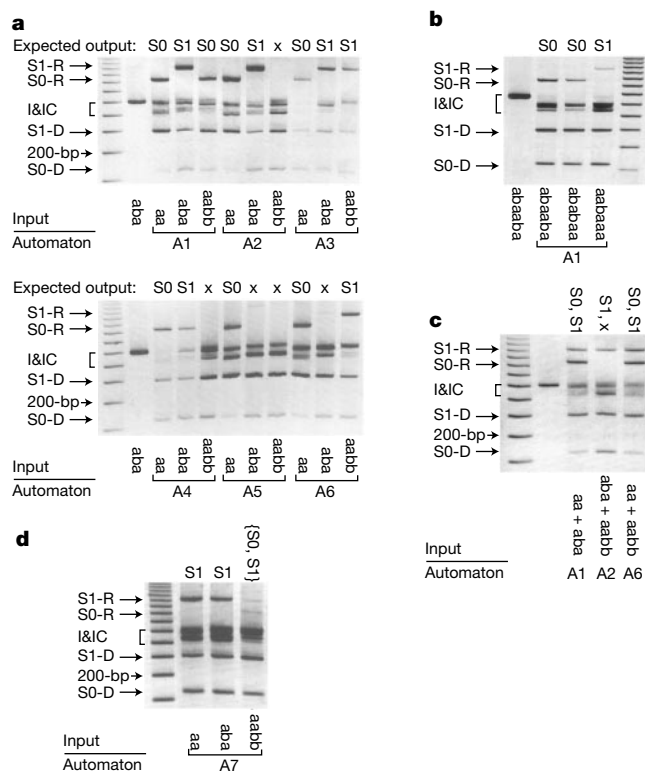


Figure 3 Running automaton programs on inputs. **a**, Experimental testing of automaton programs A1–A6 (Fig. 1a and d). The molecules in each lane are the output-detection molecules S0-D and S1-D, molecules encoding intermediate configurations capped by yet-to-be-processed input molecules and followed by output-reporting molecules S0-R or S1-R, which are missing in case the computation suspends (for example, A2, ‘at most one b’, suspends on aabb). Expected locations of different species and a size marker of 200 bp are indicated by arrows to the left. Input molecules are specified below each lane, and expected outputs (S0, S1 or suspension (x)) are indicated above each lane. Reactions with the same software are grouped in triplets, the software indicated below. I&IC, input and intermediate configurations. **b**, Computations with A1 software performed over 6-symbol-long input molecules. All inputs are of the same size, shown on a reference lane. **c**, Parallel computation. Input mixture composition and the software used are indicated below the lanes. **d**, Demonstration of a non-deterministic computation.

at least 99%, implying more than 99.8% fidelity per transition (Fig. 4c).

We have estimated the computing performance and energy consumption of the system. A computation over 2.5 pmol (1.5×10^{12} molecules) of a 4-symbol-long input rendered output-reporting molecules with ~50% yield, producing 7.5×10^{11} outputs in 4,000 s. As each output is the result of five transitions, the computing performance is of the order of 10^9 transitions per second. As for energy consumption—in each transition two ATP molecules were consumed, releasing 1.5×10^{-19} J. Multiplying this number by the transition rate provides an energy consumption rate of $\sim 10^{-10}$ J s⁻¹, or 10^{-10} W. Cost of construction of the input, software and hardware is not normally included when measuring the energy consumption of electronic computers and is not included here, although our present system will require recycling of ‘software molecules’ and of course additional energy to separate and ‘read’ the output via gel electrophoresis. This adds a significant continuing energy cost not normally encountered in electronic computers.

Any increase in the complexity of our system (number of symbols times the number of states) is bounded from above by the number of different non-palindromic sticky ends. The distribution between states and symbols depends on the length of the spacer between the

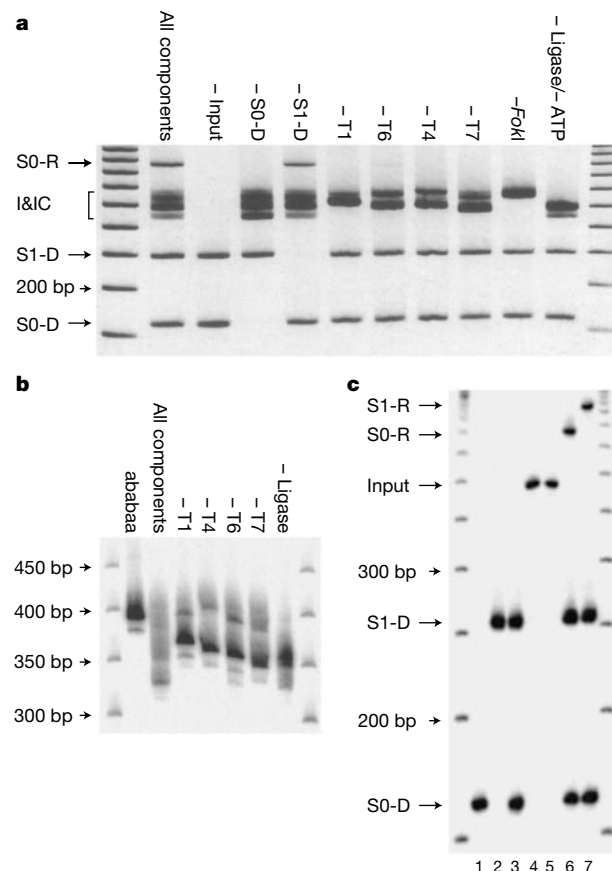


Figure 4 Verification of the operation mechanism. **a**, Identification of the essential components. Each lane is a computation reaction with one component omitted (above). The ligase or ATP-free reactions were identical and are represented by one lane. **b**, Close inspection of the reaction intermediates analysed by native PAGE (5%). Sample composition is indicated above the gel image. The ababaa lane contains labelled input. ‘All components’ lane is an output-detectors-free A1 computation, performed for reference. All other lanes lack the indicated components. The size markers are indicated to the left. **c**, An estimation of system fidelity analysed by native PAGE (6%). The composition of different lanes is as follows: 1, pure labelled S0-D; 2, pure labelled S1-D; 3, input-free computation reaction; 4, pure labelled ababaa; 5, pure labelled aabaaa; 6, a computation reaction over ababaa with unlabelled input and labelled output-detection molecules; 7, a similar reaction over aabaaa. Locations of the different components and the size markers of 200 bp and 300 bp are indicated on the left.

recognition site and the restriction site of the particular restriction enzyme employed, with *FokI* enabling a machine with at most three states and several tens of symbols. But the discovery or engineering of new class IIS enzymes with longer spacers and/or longer sticky ends²¹ might allow the construction of automata with increased complexity, linearly related to size of the spacer and exponentially related to the size of the sticky end formed on restriction. □

Methods

Synthetic DNA

Double-stranded synthetic DNA molecules were prepared by annealing 2,000 pmol of commercially obtained deoxyoligonucleotides (Sigma-Genosys) in a final volume of 10 µl of 10 mM Tris-HCl buffer, pH 8.0, containing 1 mM EDTA and 50 mM NaCl. The annealing was performed by heating the solution to 94 °C followed by slow cooling. The formation of a duplex was confirmed by native PAGE (20%). The oligomers were 5'-phosphorylated and PAGE-purified by the supplier and used without further purification.

Input molecules

These were constructed stepwise by ligating one or more synthetic DNA segments of the

desired sequence to a 1,457-bp fragment obtained by digestion of the pBluescript II SK+ plasmid (Stratagene) with *FokI*, followed by polymerase chain reaction (PCR) amplification of the coding segment and a 300- or 325-bp tail region. The sequences of the resulting input molecules were confirmed by sequencing.

Output-detecting molecules

The output-detecting molecule for the S0 output (S0-D) was formed by ligating a synthetic adapter of 30 bp containing a *FokI* recognition site to a 181-bp fragment obtained by digesting the pBluescript II SK+ plasmid with *FokI*, PCR amplification and additional *FokI* digestion to form the 160-bp fragment bearing the desired sticky end. The output-detecting molecule for the S1 output (S1-D) was obtained by PCR amplification of a 285-bp fragment corresponding to positions 1,762–2,047 of the pBluescript II SK+ plasmid followed by *FokI* digestion of the PCR product to form a 250-bp fragment.

Computation reactions

Reactions were set by mixing 2.5 pmol of the input molecule, 1.5 pmol of each output-detecting molecule and 15 pmol of each transition molecule with 12 units of *FokI* and 120 units of T4 DNA Ligase (both from New England Biolabs) in 120 μ l of NEB4 buffer supplemented with 1 mM ATP and incubating at 18 °C for 70 min. In case of multiple inputs in the same reaction, equal amounts were used, summing up to 2.5 pmol. The mixtures were purified by the Qiagen PCR purification kit and eluted using 30 μ l EB buffer (Qiagen). Aliquots (10 μ l) were assayed by gel electrophoresis using 3% MetaPhor agarose (FMC Bioproducts) unless indicated otherwise. The lengths of the DNA species were verified using a commercial 50-bp DNA step ladder (Promega). To further confirm that output reporting molecules were formed as expected, we amplified by PCR and sequenced the output-detection molecule/output molecule junction region in both output-reporting molecules S0-R and S1-R and found the expected sequences (not shown).

Received 24 January; accepted 21 September 2001.

1. Turing, A. M. On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc. II Ser.* **42**, 230–265 (1936).
2. Hopcroft, J. E., Motwani, R. & Ullman, J. D. *Introduction to Automata Theory, Languages and Computation* 2nd edn (Addison-Wesley, Boston, Massachusetts, 2000).
3. McCulloch, W. S. & Pitts, W. A logical calculus immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133 (1943).
4. Bennet, C. H. The thermodynamics of computation—a review. *Int. J. Theor. Phys.* **21**, 905–940 (1982).
5. Rothmund, P. W. K. in *DNA Based Computers: Proceedings of the DIMACS Workshop, April 4, 1995, Princeton University* (eds Lipton, R. J. & Baum, E. B.) 75–119 (American Mathematical Society, Providence, Rhode Island, 1996).
6. Smith, W. D. in *DNA Based Computers: Proceedings of the DIMACS Workshop, April 4, 1995, Princeton University* (eds Lipton, R. J. & Baum, E. B.) 121–185 (American Mathematical Society, Providence, Rhode Island, 1996).
7. Garzon, M. et al. in *Automata Implementation: Lecture Notes in Computer Science 1436* (eds Wood, D. & Yu, S.) 56–74 (Springer, Berlin, 1998).
8. Shapiro, E. & Karunaratne, K. S. G. Method and system of computing similar to a Turing machine. US Patent 6,266,569 (2001).
9. Adelman, L. M. Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021–1024 (1994).
10. Lipton, R. J. DNA solution of hard computational problem. *Science* **268**, 542–545 (1995).
11. Ouyang, Q., Kaplan, P. D., Liu, S. & Libchaber, A. DNA solution of the maximal clique problem. *Science* **278**, 446–449 (1997).
12. Landweber, L. F., Lipton, R. J. & Rabin, M. O. in *DNA Based Computers III: DIMACS Workshop, June 23–27, 1997, University of Pennsylvania* (eds Rubin, H. & Wood, D. H.) 161–172 (American Mathematical Society, Providence, Rhode Island, 1997).
13. Liu, Q. et al. DNA computing on surfaces. *Nature* **403**, 175–179 (2000).
14. Faulhammer, D., Cukras, A. R., Lipton, R. J. & Landweber, L. F. Molecular computation: RNA solutions to chess problems. *Proc. Natl Acad. Sci. USA* **97**, 1385–1389 (2000).
15. Ruben, A. J. & Landweber, L. F. The past, present and future of molecular computing. *Nature Rev. Mol. Cell Biol.* **1**, 69–72 (2000).
16. Winfree, E., Liu, F. R., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–544 (1998).
17. Sakamoto, K. et al. State transitions by molecules. *Biosystems* **52**, 81–91 (1999).
18. Hartemink, A. J., Gifford, D. K. & Khodor, J. Automated constraint-based nucleotide sequence selection for DNA computation. *Biosystems* **52**, 227–235 (1999).
19. Khodor, J. & Gifford, D. K. Design and implementation of computational systems based on programmed mutagenesis. *Biosystems* **52**, 93–97 (1999).
20. Mao, C., LaBean, T. H., Reif, J. H. & Seeman, N. C. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature* **407**, 493–496 (2000).
21. Chandrasegaran, S. & Smith, J. Chimeric restriction enzymes: What is next? *Biol. Chem.* **380**, 841–848 (1999).

Acknowledgements

We thank I. Sagi and A. Yonath for use of their laboratory for our first set of experiments, and S. Shuping, A. Regev and N. Kessler for assistance and advice. E.K. is the incumbent of the Benno Gitter and Ilana Ben-Ami Chair of Biotechnology, Technion. Z.L. is the Incumbent of The Maxwell Ellis Professorial Chair in Biomedical Research. This work was supported by the Dolphi and Lola Ebner Center for Biomedical Research.

Correspondence and requests for materials should be addressed to E.S. (e-mail: Ehud.Shapiro@weizmann.ac.il).

Superconductivity in CaCuO_2 as a result of field-effect doping

J. H. Schön*, M. Dorget†‡, F. C. Beuran†§, X. Z. Zu‡, E. Arushanov‡||, C. Deville Cavellin†‡ & M. Laguès‡

* Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Murray Hill, New Jersey 07974-0636, USA

† GPM-Université Paris, 12-61 avenue De Gaulle, 94010 Créteil Cedex, France

‡ Surfaces et Supraconducteurs CNRS UPR5-ESPCI, 10 rue Vauquelin, 75005 Paris, France

§ Wintici SA, 17 rue Jean Moulin, 94 300 Vincennes, France

|| Institute of Applied Physics, Academy of Sciences of the Moldova Republic, Academiei str. 5, Kishinev 277028, Moldova

Understanding the doping mechanisms in the simplest superconducting copper oxide—the infinite-layer compound ACuO_2 (where A is an alkaline earth metal)—is an excellent way of investigating the pairing mechanism in high-transition-temperature (T_c) superconductors more generally^{1–4}. Gate-induced modulation of the carrier concentration^{5–7} to obtain superconductivity is a powerful means of achieving such understanding: it minimizes the effects of potential scattering by impurities, and of structural modifications arising from chemical dopants. Here we report the transport properties of thin films of the infinite-layer compound CaCuO_2 using field-effect doping. At high hole- and electron-doping levels, superconductivity is induced in the nominally insulating material. Maximum values of T_c of 89 K and 34 K are observed respectively for hole- and electron-type doping of around 0.15 charge carriers per CuO_2 . We can explore the whole doping diagram of the CuO_2 plane while changing only a single electric parameter, the gate voltage.

In 1988 Siegrist *et al.* were the first to synthesize an infinite-layer copper oxide (ILCO)⁸. Because of their structural simplicity (Fig. 1), ILCOs are ideally suited as model systems providing insights into mechanisms of high- T_c superconductivity in copper oxides⁹. But the undoped compound, prepared⁸ under ambient pressure with the composition $\text{Ca}_{0.86}\text{Sr}_{0.14}\text{CuO}_2$, exhibits only antiferromagnetic and no superconducting properties¹⁰. Later, n-type doping was obtained with the bulk compound $\text{Sr}_{1-x}\text{Ln}_x\text{CuO}_2$ ($x \leq 0.1$ –0.15) synthesized under a pressure of 3–6 GPa (refs 1, 11–14). In addition, electron-doped, pseudomorphically stabilized thin films were also grown using various techniques^{15–17}. Superconducting transitions were observed in these doped compounds with a maximum T_c

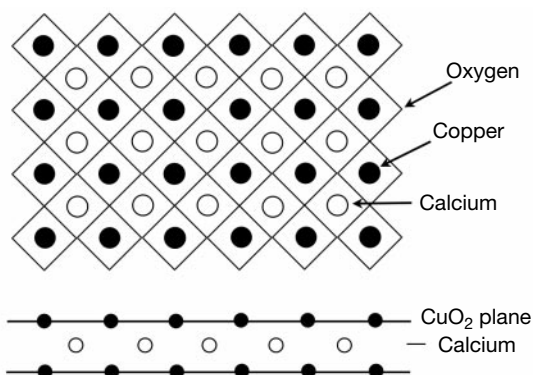


Figure 1 Schematic structure of the infinite-layer compound CaCuO_2 . This is the 'parent structure' of the layered high-temperature superconductors. The structure is tetragonal, and consists of CuO_2 planes and cationic Ca planes, alternately stacked along the c -axis.

of 44 K at $x = 0.05\text{--}0.1$) (refs 1, 11–14). But researchers were unable to control chemical substitution to study doping models.

Bulk $(\text{Sr}_{1-x}\text{Ca}_x)_{1-y}\text{CuO}_2$ and $(\text{Sr}_{1-x}\text{Ba}_x)_{1-y}\text{CuO}_2$ solid solutions, synthesized under high pressure^{18–20} with a maximum T_c of 110 K, were reported to be ILCOs with p-type doping. It was proposed that the superconducting properties in these p-type compounds are related to the presence of defect layers^{2,4}, or to the presence of impurity phases such as $(\text{Sr,Ca})_3\text{Cu}_2\text{O}_x$ (ref. 19) or tetragonal $\text{Sr}_{n+1}\text{Cu}_n\text{O}_{2n+1+\delta}$ compounds²⁰. These defects probably act as local charge reservoir layers for adjacent, unperturbed CuO_2 sheets, changing four-fold Cu–O coordination to five-fold as well²¹. The presence of apical oxygen in the superconducting p-type bulk compounds was thus suspected²². Moreover, as in the case of n-type doping, the onset T_c of 110 K does not depend significantly on the composition, indicating that the superconducting properties are due to a secondary phase rather than to the infinite-layer structure⁴. Sr-rich ILCOs were studied mainly because of the thermodynamic metastability of pure Ca compounds. CaCuO_2 films^{4,17,23} and single crystals²⁴ have been prepared and characterized, but no superconductivity was observed. Superconductivity has been observed in $\text{SrCuO}_2/\text{BaCuO}_2$ (ref. 25) as well as in $\text{BaCuO}_2/\text{CaCuO}_2$ superlattices²⁶ with a maximum T_c value of about 80 K. But the Ba-based ILCO tends to be very unstable. It can easily include excess oxygen ($\text{BaCuO}_{2+\delta}$), thereby acting as a charge reservoir block for the second constituent ILCO layer ($\text{Ca}_{1-x}\text{Sr}_x$) CuO_2 (ref. 26).

Hence, various questions arise concerning the doping-dependence of superconductivity in ILCOs and their relation to defects, structural changes or impurity phases. Compared to chemical doping, field-effect doping^{5–7} is not limited by dopant solubility, compositional stability, structural transitions or changes in defect chemistry. This technique allows the investigation of properties of a given material as a function of the doping level without inducing disorder or additional defects. Consequently, the changes in material properties are directly related to the effect of the carrier concentration. There is a possibility of exploring the whole phase diagram of the CuO_2 plane using a single electric control parameter, the applied gate voltage. But it might be difficult to combine the field-effect doping technique of single CuO_2 layers with other useful probes, such as angle-resolved photoemission spectroscopy or magnetic measurements.

We grew single-crystalline 30-nm-thick films of CaCuO_2 by molecular beam epitaxy (MBE) on a SrTiO_3 (100) substrate¹⁷.

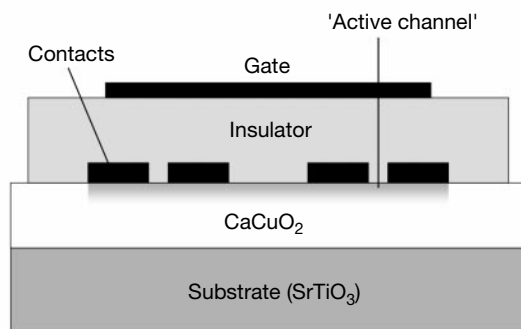


Figure 2 Structure of the field-effect device. The CaCuO_2 film is grown by molecular beam epitaxy (MBE) on a SrTiO_3 (100) substrate. The growth was performed under atomic oxygen at a substrate temperature of 600 °C. The crystal structure and composition were determined using four-circle X-ray diffraction, high-resolution transmission electron microscopy and Rutherford backscattering, revealing a high quality of the MBE-grown films. No signs of apical oxygen have been observed for these ILCO films. Electrical contacts are prepared on top of the film by Pt deposition. A gate dielectric is provided by sputtering an insulating layer of Al_2O_3 . Finally, an evaporated gold layer acts as a gate electrode.

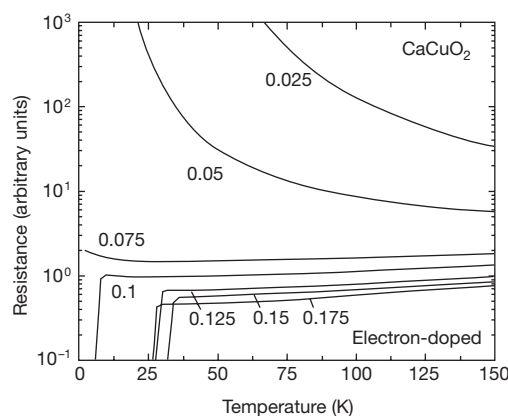


Figure 3 Resistance of a CaCuO_2 film as a function of temperature and electron-doping level x . Transitions from insulating to metallic and finally to superconducting behaviour are observed. From the suppression of superconductivity in a magnetic field, an in-plane coherence length of 15 to 12 Å is estimated for the optimally doped ILCO, which is in good agreement with values reported for other copper oxide superconductors.

Field-effect devices were prepared on top of these films (Fig. 2). By applying a negative or positive voltage to the gate, the doping level in the active channel at the insulator/ CaCuO_2 interface can be modulated. Electric-field modulation of the properties of ILCO layers has also been reported by other groups^{27,28}. The observation of ambipolar transistor action, that is, accumulation of electrons or holes depending on the sign of the gate voltage, demonstrates the high quality of the interface²⁹.

As-grown, undoped CaCuO_2 thin films are insulating, exhibiting a thermally activated resistivity (ρ) (ref. 17). With increasing gate voltage, the conductance increases and the temperature dependence fits well with the variable-range hopping mechanism, and at even higher bias a power-law dependence is observed^{4,17}. Figures 3 and 4 show the resistance as a function of temperature for electron- and hole-doping levels exceeding $x = 0.025$, where x is the number of charge carriers per CuO_2 unit. At hole- and electron-doping levels higher than $x = 0.035$ and $x = 0.08$ superconductivity is observed. The superconducting phase diagram of gate-doped CaCuO_2 is shown in Fig. 5. This diagram exhibits apparent electron and hole superconductivity, and bears a strong resemblance to results

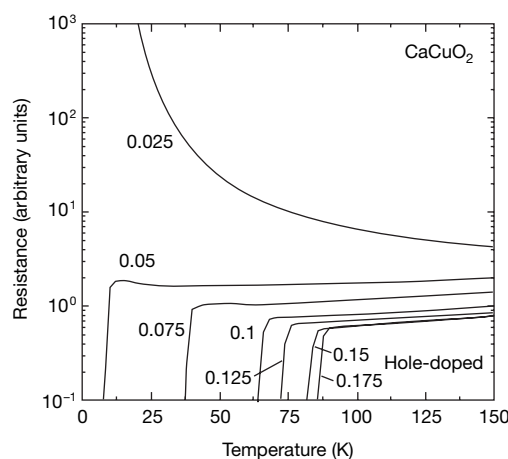


Figure 4 Resistivity of a CaCuO_2 film as a function of temperature and hole-doping level x . A transition from insulating to metallic and, finally, to superconducting behaviour is observed.

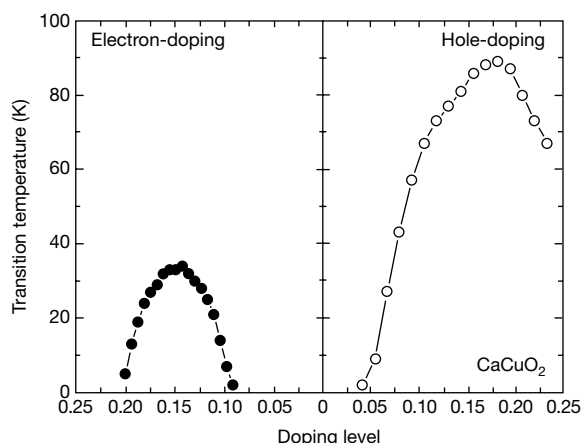


Figure 5 The superconducting phase diagram of CaCuO_2 . A maximum T_c of 89 and 34 K is obtained at optimal hole and electron doping, respectively. Doping level x is defined as number of charge carriers per CuO_2 unit.

observed for other copper oxides³⁰. A maximum T_c of 89 and 34 K is obtained for optimally hole- and electron-doped ($x \approx 0.15$ – 0.17) CaCuO_2 , respectively (see Fig. 5). A slight dip in T_c as a function of x is observed at $x \approx 1/8$. For comparison, the temperature dependence of the normalized resistance in the n-type underdoped, optimally doped and overdoped regime is plotted in Fig. 6. In the underdoped regime a 'convex', and in the overdoped regime a 'concave', behaviour of the normal-state resistance is observed. (Concave and convex mean that the second derivative of the resistance as a function of temperature is smaller or larger than zero, respectively.) In the case of optimum doping, a linear temperature dependence of ρ is measured. In addition, at low doping levels ($x \approx 0.1$) a logarithmic divergence of ρ is found. These observations for n- as well as p-type doping of CaCuO_2 are in accordance with non-Fermi-liquid-like behaviour for underdoped samples and Fermi-liquid-like metallic transport in the overdoped regime³⁰. More detailed studies of the normal-state properties might lead to a better understanding of transport and superconductivity in ILCOs and copper oxides in general.

The observation of superconductivity in the same CaCuO_2 film for both hole and electron doping indicates that superconductivity is not related to defects or secondary phases, as it is very unlikely that such structural defects could be responsible for both types of superconductivity. In addition, different defects have been supposed to be responsible for n- and p-type superconductivity in ILCOs⁴. Moreover, the results reveal that when using field-effect

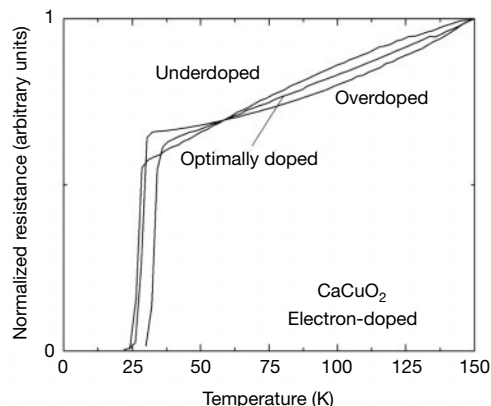


Figure 6 Normalized resistance as a function of temperature for n-type under-doped, optimally doped and overdoped CaCuO_2 . See text for discussion.

doping, no apical oxygen is necessary for hole-type superconductivity with transition temperatures close to 100 K, which has caused controversy in the literature^{3,4}. Our results indicate the intrinsic doping properties of the CuO_2 plane. The possibility of tuning the properties of copper oxide materials over a wide doping range can be seen as an ideal starting point to elucidate the complex physics underlying this class of materials, which in turn might lead to optimization of the superconducting properties. Moreover, this technique also provides ways to investigate doping dependence in other correlated insulator systems. □

Received 30 July; accepted 12 October 2001.

- Smith, M. G., Manthiram, A., Zhou, J., Goodenough, J. B. & Markert, J. T. Electron-doped superconductivity at 40 K in the infinite-layer compound $\text{Sr}_{1-x}\text{Nd}_x\text{CuO}_2$. *Nature* **351**, 549–551 (1991).
- Azuma, M., Hiroi, Z., Takano, M., Bando, Y. & Takeda, Y. Superconductivity at 110 K in the infinite-layer compound $(\text{Sr}_{1-x}\text{Ca}_x)_{1-y}\text{CuO}_2$. *Nature* **356**, 775–776 (1992).
- Zhang, H. *et al.* Identity of planar defects in the infinite-layer copper oxide superconductor. *Nature* **370**, 352–354 (1994).
- Laguès, M. *et al.* in *Coherence in High Temperature Superconductors* (eds Deutscher, G. & Revcoleschi, A.) 70–98 (World Scientific, Singapore, 1996).
- Schön, J. H., Kloc, Ch., Haddon, R. C. & Batlogg, B. A superconducting field-effect switch. *Science* **288**, 656–658 (2000).
- Schön, J. H., Kloc, Ch. & Batlogg, B. Superconducting at 52 K in hole-doped CaO . *Nature* **408**, 549–552 (2000).
- Ahn, C. H. *et al.* Electrostatic modulation of superconductivity in ultrathin $\text{GdBa}_2\text{Cu}_3\text{O}_{7-x}$ films. *Science* **294**, 1152–1155 (1999).
- Siegrist, T., Zahurak, S. M., Murphy, D. W. & Roth, R. S. The parent structure of the layered high-temperature superconductors. *Nature* **334**, 231–232 (1988).
- Gupta, A. Growth and properties of thin films of the doped infinite-layer compounds $\text{Ca}_{1-x}\text{Ln}_x\text{CuO}_2$ ($\text{Ln} = \text{Sm}, \text{Nd}, \text{Y}$). *Physica C* **231**, 389–394 (1994).
- Lombardi, A., Mali, M., Roos, J., Brinkman, D. & Mangelschots, I. Temperature dependence of the sublattice magnetization of the antiferromagnet $\text{Ca}_{0.85}\text{Sr}_{0.15}\text{CuO}_2$. *Phys. Rev. B* **54**, 93–96 (1996).
- Er, G., Miyamoto, Y., Kanamuru, F. & Kikkawa, S. Superconductivity in the infinite-layer compound $\text{Sr}_{1-x}\text{La}_x\text{CuO}_2$ prepared under high pressure. *Physica C* **181**, 206–208 (1991).
- Wooten, C. L. *et al.* The pressure dependence of T_c in the infinite-layer electron-doped compound $\text{Sr}_{0.84}\text{Nd}_{0.16}\text{CuO}_2$. *Physica C* **192**, 13–17 (1992).
- Korczak, W., Peroux, M. & Strobel, P. Superconductivity in $\text{Sr}_{0.85}\text{La}_{0.15}\text{CuO}_2$. *Physica C* **193**, 303–308 (1992).
- Ikedo, N., Hiroi, Z., Azuma, M., Takano, M. & Bando, Y. Synthesis and superconducting properties of the infinite-layer compounds $\text{Sr}_{1-x}\text{Ln}_x\text{CuO}_2$ ($\text{Ln} = \text{La}, \text{Nd}, \text{Sm}, \text{Gd}$). *Physica C* **210**, 367–372 (1993).
- Niu, C. & Lieber, C. M. Growth of the infinite layer phase of $\text{Sr}_{1-x}\text{Nd}_x\text{CuO}_2$ by laser ablation. *Appl. Phys. Lett.* **61**, 1712–1714 (1992).
- Adachi, H., Satoh, T., Ishikawa, Y., Setsune, K. & Wase, K. Superconducting $(\text{Sr}, \text{Nd})\text{CuO}_2$ thin films with infinite-layer structure. *Physica C* **196**, 14–16 (1992).
- Xie, X. M. *et al.* Transport mechanisms in infinite layer compounds grown by molecular beam epitaxy. *Appl. Phys. Lett.* **67**, 1671–1673 (1995).
- Hiroi, Z., Azuma, M., Takano, M. & Takeda, Y. Structure and superconductivity of the infinite-layer compound $(\text{Ca}_{1-x}\text{Sr}_x)_{1-y}\text{CuO}_{2-y}$. *Physica C* **208**, 286–296 (1993).
- Prouteau, C., Strobel, P., Capponi, J. J., Chailout, C. & Tholence, J. L. Optimization of superconductivity in the high-pressure Sr-Ca-Cu-O system. *Physica C* **228**, 63–72 (1994).
- Shaked, H. *et al.* Superconductivity in the Sr-Ca-Cu-O system and the phase with infinite-layer structure. *Phys. Rev. B* **51**, 11784–11790 (1995).
- Feenstra, R. *et al.* Electron-doped and hole-doped infinite layer $\text{Sr}_{1-x}\text{Ca}_x\text{CuO}_{2-\delta}$ films grown by laser molecular beam epitaxy. *Physica C* **224**, 300–316 (1994).
- Zhou, X., Yao, Y. S., Jia, S. J., Dong, C. & Zhao, Z. Stability and doping of infinite-layer compound $(\text{Ca}_{1-x}\text{Sr}_x)\text{CuO}_2$ at ambient pressure. *J. Mater. Sci.* **30**, 952–954 (1995).
- Vallionis, A., Brazdeikis, A. & Flodström, A. S. Observation of local oxygen displacements in CuO_2 planes induced by a misfit strain in heteroepitaxially grown infinite-layer-structure $(\text{Ca}_{1-x}\text{Sr}_x)\text{CuO}_2$ films. *Phys. Rev. B* **55**, R6152–R6155 (1997).
- Kopin, E. M. *et al.* $\text{Ca}_{1-x}\text{R}_x\text{CuO}_2$ ($\text{R} = \text{Sr}, \text{La}$) single crystals with infinite-layer structure: high Ar gas pressure synthesis and properties. *Physica C* **282**–**287**, 483–484 (1997).
- Norton, D. P. *et al.* Superconductivity in SrCuO_2 - BaCuO_2 superlattices: formation of artificially layered superconducting materials. *Science* **265**, 2074–2077 (1994).
- Balestrino, G., Martellucci, S., Medaglia, P. G., Paoletti, A. & Petrocelli, G. Dependence of the critical temperature on n in $(\text{BaCuO}_2)_2/(\text{CaCuO}_2)_n$ superlattices. *Phys. Rev. B* **58**, R8925–R8928 (1998).
- Ahn, C. H. *et al.* Ferroelectric field effect in epitaxial thin film oxide $\text{SrCuO}_2/\text{Pb}(\text{Zr}_{0.52}\text{Ti}_{0.48})\text{O}_3$ heterostructures. *Science* **269**, 373–376 (1995).
- Kawahara, T., Sugiuchi, N., Komai, E., Terashima, T. & Bando, Y. Hole and electron doping in ultrathin films with the infinite layer structure by the electric field effect. *Physica C* **276**, 127–131 (1997).
- Schön, J. H., Berg, S., Kloc, Ch. & Batlogg, B. Ambipolar pentacene field-effect transistors and inverters. *Science* **287**, 1022–1024 (2000).
- Maple, M. B. High-temperature superconductivity. *J. Magn. Magn. Mater.* **177**–**181**, 18–30 (1998).

Acknowledgements

We thank G. Blumberg, Ch. Kloc, H. Y. Hwang and C. M. Varma for discussions.

Correspondence and requests for materials should be addressed to J.H.S. (e-mail: hendrik@lucent.com).

Increased damage from fires in logged forests during droughts caused by El Niño

F. Siegert^{*†}, G. Ruecker[‡], A. Hinrichs[§] & A. A. Hoffmann^{||}

^{*} Ludwig Maximilians University, Department of Biology, Luisenstrasse 14, 80333 München, Germany

[†] Remote Sensing Solutions GmbH, Würthstrasse 49, 81667 München, Germany

[‡] ZEBRIS GIS and Consulting, Lipowskystrasse 26, 81373 München, Germany

[§] Sustainable Forest Management Project (SEMP-GTZ-MoFEC), PO Box 1087, Samarinda 75123, Kalimantan Timur, Indonesia

^{||} Integrated Forest Fire Management Project IFFM/GTZ, Pekantoran Dinas Kehutanan, Jln. Harmonika, Samarinda, 75001, Kalimantan Timur, Indonesia

In 1997–98, fires associated with an exceptional drought caused by the El Niño/Southern Oscillation (ENSO) devastated large areas of tropical rain forests worldwide. Evidence suggests that in tropical rainforest environments selective logging may lead to an increased susceptibility of forests to fire^{1–4}. We investigated whether this was true in the Indonesian fires, the largest fire disaster ever observed^{5,6}. We performed a multiscale analysis using coarse- and high-resolution optical and radar satellite imagery assisted by ground and aerial surveys to assess the extent of the fire-damaged area and the effect on vegetation in East Kalimantan on the island of Borneo. A total of 5.2 ± 0.3 million hectares including 2.6 million hectares of forest was burned with varying degrees of damage. Forest fires primarily affected recently logged forests; primary forests or those logged long ago were less affected. These results support the hypothesis of positive feedback between logging and fire occurrence⁴. The fires severely damaged the remaining forests and significantly increased the risk of recurrent fire disasters by leaving huge amounts of dead flammable wood.

During the last three decades pressure on forests by logging and massive transmigration has strongly increased on the island of Borneo⁷. According to official sources⁸, more than 180 million cubic metres of large logs have been harvested from the forests of East Kalimantan since 1969. Fires were often started for large-scale forest conversion into plantations and land clearance for agriculture⁶. Fires also occurred because impoverished people used fire for hunting and for the collection of turtles from swamp forests and in disputes over land ownership⁹. At the end of the ENSO episode in the early spring of 1998 fires flared up in East Kalimantan, an area that was not much affected in 1997 (refs 6 and 10). Fires associated with extended ENSO-related droughts were described as early as 1914 for East Kalimantan but these were small in scale³. Undisturbed tropical rainforest is normally highly resistant to fire because of low loads of available fuel, low fuel-energy content and high humidity even during drought^{11–13}. Fire became a threat to the rainforests in east Kalimantan only recently. Although the drought of 1971–72 did not result in significant fire damage, 3.5 million hectares of mainly forested land were destroyed during the 1982–83 ENSO, the largest fire disaster observed in tropical rainforests until then^{14,15}.

Fire occurrence in 1997–98 was monitored using coarse-resolution images from the advanced very-high-resolution radiometer (AVHRR, 1,100-m ground resolution) aboard the National Oceanic and Atmospheric Administration (NOAA) 12 and NOAA 14 satellites, which were processed to detect 'hot spots'. During a period of 290 days from August 1997 to May 1998 more than 65,000 hot spots were recorded in East Kalimantan. The time series of hot spots showed that the fires started in the Mahakam river basin, which had already been severely affected by the 1982–83 fires. Subsequently

fires spread to other regions and propagated often as a fire front (red hot spots in Fig. 1a).

Assessment of fire impact requires satellite images with higher resolution than that of NOAA-AVHRR. Optical satellite systems were severely hampered during the fires by haze and rain following the drought. Therefore we used data acquired by the high-resolution active microwave instrument (AMI, 25-m ground resolution), a synthetic aperture radar (SAR) onboard the European Remote Sensing Satellite (ERS)-2 to locate and assess the extent of the burned area. Active microwave systems are able to penetrate clouds and haze. Difference detection techniques applied to pairs of ERS-2 SAR images allowed us to map the burned area with high accuracy and to assess the level of fire damage from changes in image intensity and texture¹⁶. The AMI instrument detects structural features of the earth's surface (volume scattering) and the moisture content of the vegetation (dielectric properties)¹⁷. Fire decreases both. Volume scattering decreases because fire consumes vegetation and moisture content decreases because fire-damaged plants lose their foliage. Furthermore, the opened canopy and the reduced leaf biomass allow more microwave backscatter from the exposed ground surface.

The total area affected by fire in East Kalimantan was 5.2 ± 0.3 million hectares (Fig. 1b). The radar analysis showed that 24% of the burned area had moderate fire damage (25–50% of trees dead), 42% had severe fire damage (50–80% of trees dead), and 34% showed total fire damage (>80% of vegetation killed). In this last category we distinguished two cases: (1) the vegetation was almost totally consumed by the fires (mostly grass and bush land); and (2) most of the standing biomass was not consumed but tree mortality was almost 100%. The latter case occurred mostly in pristine peat swamp forests (12% of the total burned area). To determine what type of vegetation was burned we produced a vegetation map differentiating six land-cover classes using ERS-2 radar images acquired in August 1997 just before the fires (Fig. 1c). The burned area was then intersected with this land-cover map (Table 1). The result shows that fire mainly affected lowland dipterocarp, secondary forests and peat swamp forests, whereas the vast majority of mangrove and highland dipterocarp forests in mountainous areas escaped the fires.

To investigate whether intensified use of forest increased fire susceptibility, all available planning data on pre-fire legal land status, such as natural forest concessions, plantations, industrial timber crops and protected forests, were collected from official sources, digitized and transferred into a geographical information system (GIS). The intersection of the radar-based fire-impact map with land-use status boundaries (Table 2) showed that the bulk of the burned area (5.2 million hectares) were timber concessions, plantations or land of undefined status (mostly agriculture, shifting cultivation and fallow land), while only 0.4 million hectares were protected (presumably pristine) forests. The highest damage occurred in pulp-wood plantations: almost two-thirds were destroyed by the fires (severe to total fire damage). Some 24% of the forest concession area (9.7 million hectares) in East

Table 1 Damaged area for different land covers

Land cover	Area (ha)	Burned (ha)	Burned (%)
Grassland (mainly <i>Imperata cylindrica</i>), low bushes	368,900	292,600	79.3
Lowland dipterocarp forest	5,379,600	2,177,900	40.5
Mangrove forest	1,042,100	91,700	8.8
Peat swamp forest	426,100	311,100	73.0
Secondary forest, plantation, farmland	2,283,400	1,723,400	75.5
Wetlands	358,700	290,400	81.0
Land cover not mapped by ERS (mainly highland dipterocarp forest)	3,882,600	330,800	8.5
Total	13,741,400	5,217,900	—

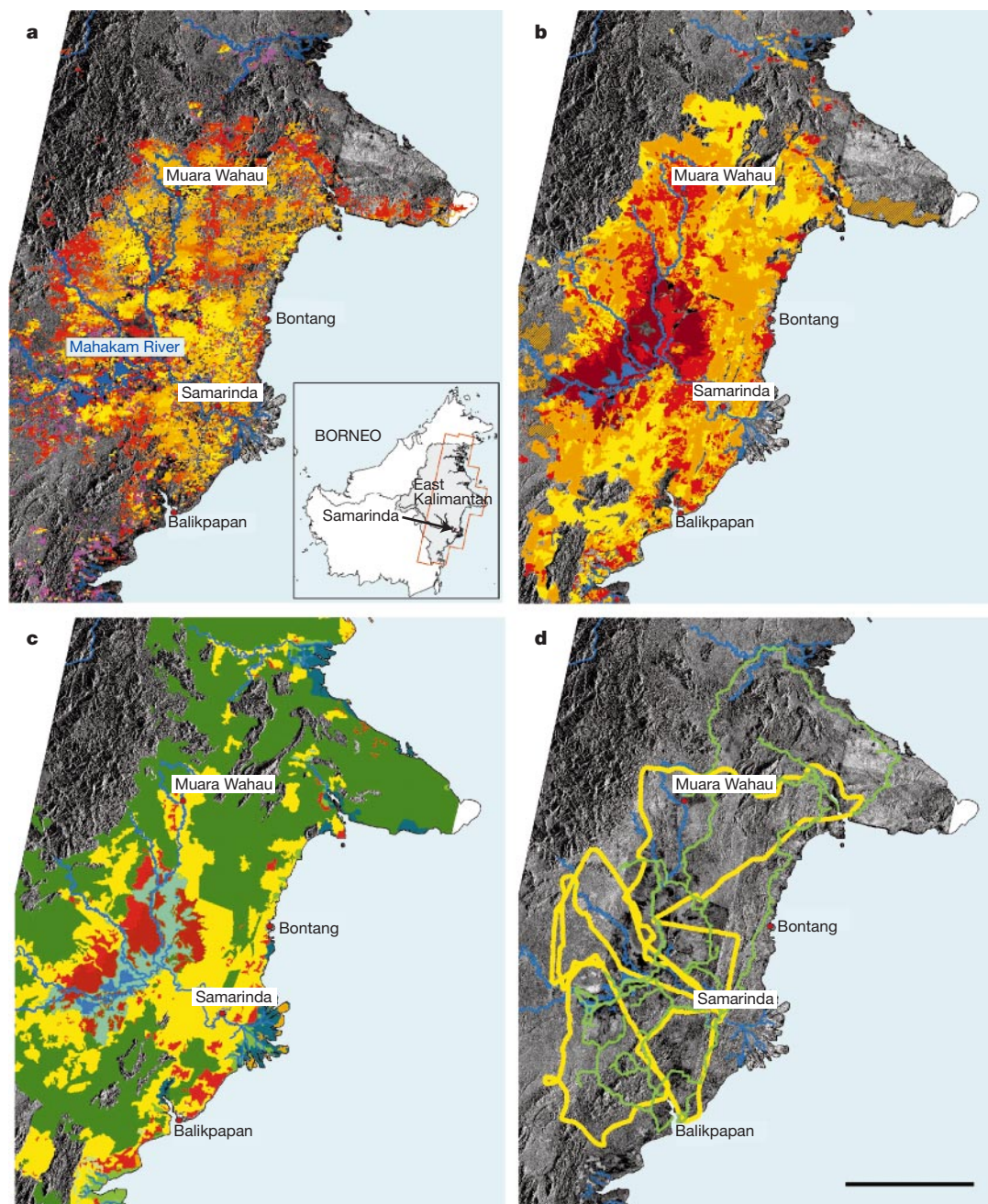


Figure 1 Fire occurrence and fire impact in 1997/98 in East Kalimantan, Borneo, Indonesia. Rivers are shown in blue. **a**, Time series of NOAA–AVHRR fire hot spots. Yellow–red colour gradient, January to April 1998; purple, August to November 1997. In 1998 the fires started in the densely populated Mahakam basin. Background is the ERS SAR mosaic, August 1997. Scale bar, 100 km. The inset shows a map of Borneo, East Kalimantan and the studied area (red outline). **b**, Fire scars and fire impact map. Damage levels are as follows: yellow, moderate; orange, severe; red, total damage of vegetation;

brown, total damage of trees in peat swamp forests (see text). **c**, Pre-fire vegetation map derived from ERS SAR images. Green, pristine and logged dipterocarp forest; brown, peat swamp forest; light green, mangrove; yellow, degraded/secondary forests/plantations; red, grassland, bushland, clearings; and aqua, wetlands. No colouring, mountainous areas (mainly highland dipterocarp forest). **d**, GPS recorded aerial (yellow) and ground surveys (green).

Table 2 Damaged area for different land uses

Land status (1998)	Area (ha)	Burned (ha)	Burned (%)	Damage (% of total burned area)		
				Moderate	Severe	Total
Forest concessions	9,771,400	2,350,000	24.0	32.7	52.6	14.7
Pulpwood plantations	1,393,100	884,000	63.4	23.7	48.6	27.7
Oil palm plantations	746,600	382,500	51.2	21.9	51.8	26.3
Protected forests	4,562,000	440,100	9.6	19.1	59.9	21.0
Other land use (undefined & agriculture)	3,275,480	1,161,300	35.4	9.2	6.0	84.8
Total (whole province)	19,748,500	5,217,900	26.4	24	42	34

Plantations include established and planned concessions. For damage classes, see text.

Kalimantan was burned and more than half of this area showed severe damage.

As official land-use data may not fully reflect the actual condition of the forest just before the disaster, we used the only available, relatively cloud-free high-resolution (30 m) Landsat-5 Thematic Mapper (TM) images (covering about 15% of the ERS study area) as an additional source of information to identify logged and undisturbed forests. A historical and a recent land-cover map were produced from images acquired in 1990/92 and during the fires in February–March 1998. To investigate fire impacts on undisturbed and logged forests the TM land-cover map was intersected with the ERS–SAR fire impact map (total matching area 3,587,100 ha). Only 5.7% of undisturbed forests were affected by fire, compared to 59% of logged forest and 70.7% of the non-forest area (Fig. 2). Fire impact was also more severe in logged forests: 48% had severe or total damage, in contrast to only 4% of the undisturbed forests. Peat swamp forests were excluded from this analysis because logging did not occur and fire behaviour was different.

A detailed field survey of fire impact was conducted in a 100,000-ha forest concession. The concession management had not implemented any operational fire protection measures. Information about the standing volume of living and dead trees and the damage level was collected by surveying the entire concession area¹⁸. The degree of fire impact was strongly correlated with the time elapsed after logging (Fig. 3). In forest areas logged between 1996–98 the volume of dead trees was almost equal to the volume of living trees, while in forest areas logged between 1969–81 the volume of living trees was almost six times higher than that of dead trees. Damage levels showed the same pattern: in recently logged forests (1996–98) severe damage was found in 49.5% of the burned area compared to only 26.3% in old logged forests and 17.3% in pristine forest (data not shown). In pristine forests and forests logged long ago dead biomass consisted mainly of litter on the forest floor causing relatively weak surface fires (onsite observations). Where logging activity had occurred, logging waste and dense undergrowth of fast-growing pioneer species (trees less than 20 cm diameter at breast height, d.b.h.) provided huge fuel loads. Open forest canopy in logged forests also increases photon flux density resulting in increased flammability¹⁹. Fires were occasionally observed reaching into the crowns of older trees with hollow trunks.

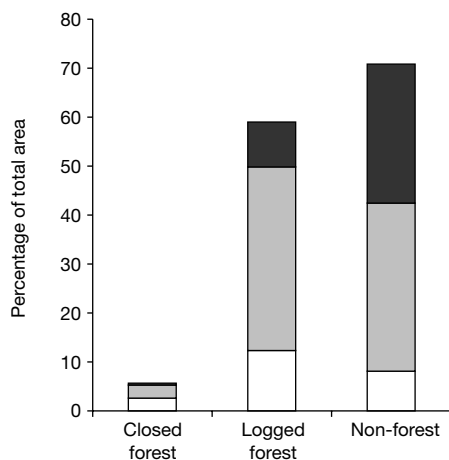


Figure 2 ERS SAR-derived fire damage for different land cover, derived from TM as a percentage of the total area of each land cover. White, moderate fire damage; grey, severe damage; black, total damage. The non-forest class includes *Imperata cylindrica* grasslands, forest mosaics, recently established plantations and agriculture, including fallow land.

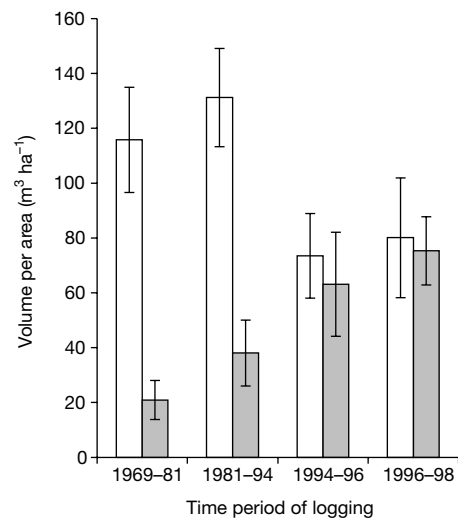


Figure 3 Fire impact measured as volume of living and dead trees in burned areas logged at different times in the past. White bars, living trees; grey bars, dead trees. The volume was measured for all trees over 20 cm d.b.h. (3,044 trees) on 306 sample plots. Error bars are 95% level of confidence for mean. Logged area (n = number of sample plots): in 1969–1980/81, 6,722 ha (n = 62); in 1981/82–1993/94, 5,464 ha (n = 51); in 1994/95–1995/96, 5,677 ha (n = 55); in 1996/97–1997/98, 7,579 ha (n = 70). Differences in percentage of dead tree volume were significant at the 0.05% level of confidence for plots logged between 1969 and 1994 and plots logged between 1994 and 1998 (one-way analysis of variance with posterior Scheffé test for group differences).

We used logging roads as an indicator of forest disturbance and prepared maps of the historical (before 1992) and recent network of logging roads (roads established between 1992 and 1998) from the TM images. Within the studied concession area we analysed fire impact (damage levels) within 1,000-m-wide strips centred on logging roads. Fire impact was much higher in proximity to recently established logging roads (65% of the area burned) compared to old logging roads (16% of the area burned) (Fig. 4).

The ground survey of one concession showed that fire destroyed

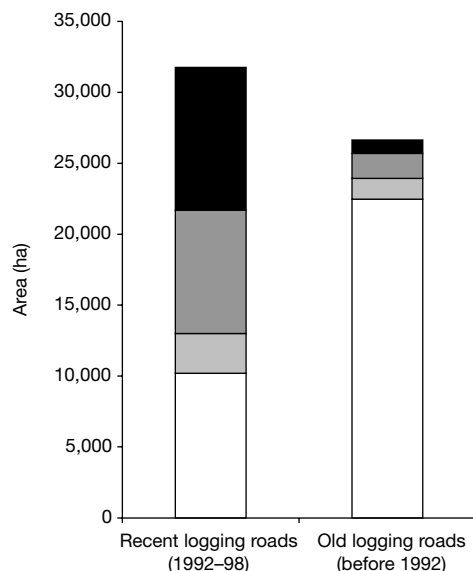


Figure 4 Fire impact within 1,000 m of logging roads. White, unburned; light grey, moderate fire damage; dark grey, severe damage; black, total damage.

an average 15 m^3 of merchantable timber per hectare in logged forests, which constitute the bulk of the burned forests. Assuming that a log volume value of 10 m^3 per hectare holds for all timber concessions affected by fire, we conservatively estimate economic losses of merchantable logs totalling more than 2 billion US dollars for the whole province of East Kalimantan (compare Table 2; the average market value of timber in Samarinda, the largest timber wholesale site, is 85 US dollars per cubic metre).

Many aspects of the 1998 fires in East Kalimantan closely resembled the fires associated with the 1982–83 ENSO event^{14,15}. However, the 1997–98 fires by far surpassed the 1982–83 disaster and occurred in most of the areas damaged previously. The area affected by fire was 40% larger and fires predominantly spread into recently opened up (that is, logged) forests in that area.

Our results indicate that recently logged forests were hit harder by fire than undisturbed or partially recovered forests. These results support the hypothesis of positive feedback between logging and fire occurrence as has also been observed in the Amazon rainforests⁴. Thus selective logging directly contributed to the unprecedented extent of the 1998 fires. Many areas burned in 1982–83 fires suffered from recurrent fires in the 1990s and did not recover into fire-resistant tropical rainforest³. Similarly, for forested areas burned in 1998 the fire hazard has greatly increased, especially where fire left behind an opened-up forest or large volumes of unburned biomass as in peat swamp forests. Unless land-use policies are changed to control logging and to introduce reduced-impact logging techniques, recurrent fires will lead to a complete loss of Borneo's lowland rainforests^{19–21}. □

Methods

Radar fire map

In all, 23 ERS-2 SAR precision images acquired in August 1997 represented the pre-fire vegetation status; 23 images acquired in July 1998 represented the post-fire situation. In addition one orbit (8 images) acquired during active burning in April 1998 was evaluated. pre- and post-fire images were co-registered to form bi-temporal image pairs consisting of the August–July and August–April orbits respectively. Co-registration was done automatically using the SAR Toolbox (ESA) with a registration error of less than one pixel. The images were calibrated to represent radar image backscatter, sub-sampled to 25-m pixel size, speckle-filtered, georeferenced (using orbital information and over 100 global positioning system (GPS) measured ground control points) and combined into a mosaic. The bi-temporal mosaic was visually interpreted onscreen using a Geographic Information System (scale 1:200,000, minimal mapping unit 100 ha). The interpretation of SAR signatures was based on GPS mapped ground observations made at 138 sites and evaluation of 42 video sequences scattered throughout the burned area along the tracks shown in Fig. 1d. Three different damage classes were defined on the basis of the estimated percentage of dead trees. NOAA–AVHRR hot-spot data served as an additional source of information during the interpretation process to reduce drought-related errors. An area was scored as 'burned' only when there was: (1) a clear decrease in backscatter or (2) a weak decrease in backscatter in conjunction with NOAA–AVHRR hot spots. Classification results were verified using more than 25 hours of GPS synchronized video material from four ground surveys and five aerial surveys conducted between April 1998 and September 1999 (Fig. 1d). The overall classification accuracy for the discrimination of burned from unburned surfaces was more than 95% based on 172 randomly selected digital video images from the aerial surveys. For discrimination of damage classes the error of commission was 10% and the error of omission was 30% for all classes. Most undetected areas belonged to damage class 25–50%. This, together with ground observations, suggests that low-intensity surface fires escape detection by both NOAA–AVHRR and ERS SAR.

Vegetation mapping

Landsat TM: Images were corrected for atmospheric moisture, georeferenced using a set of GPS measured ground control points and visually interpreted onscreen using a classification scheme based on ground observations (scale 1:100,000, minimal mapping unit 50 ha). ERS radar: The land-cover map was produced by visual interpretation of speckle-filtered, texture-enhanced artificial RGB (red–green–blue) images²² (scale 1:200,000, minimal mapping unit 150 ha) with the following classes: lowland dipterocarp forest (pristine and logged), secondary forests including plantations and farmland, mangrove and peat swamp forests (generally pristine because the amount of merchantable timber is low), grass- and wetlands. The TM mosaic was co-registered onto the ERS mosaic using 50 ground control points (in flat areas) resulting in an accuracy of r.m.s. ± 0.86 pixel (root mean square error, using a second-order polynomial transform). Overall agreement of the ERS and TM interpretation was 71% and 74% for forest/non-forest discrimination.

Ground survey in timber concession

The concession selected for the ground survey is a model forest area of the Ministry of Forestry (MoF), which is supported by the German government (technical support given by GTZ). The area was chosen because all activities before and during the fires are well documented and the MoF permitted us to conduct the survey in coordination with the concession management. The survey method of determining damage level and the standing volume of living and dead trees was a systematic line survey along 53 north–south oriented survey lines (spaced 1.2 km apart, 4–40 km long) covering the entire concession area (approximately 100,000 ha). Fire damage level (moderate, severe or total) was continuously estimated to the left and to the right side of each survey line and the average for each side separately recorded for 300-m sections. The average viewing distance from the survey lines into the forest was 60 m depending on topography and fire damage level. Every 900 m on every survey line a $20\text{ m} \times 125\text{ m}$ sampling plot (0.25 ha) was established and all standing living and dead trees above 20 cm d.b.h. were measured and recorded. The forest inventory was performed from 18 June to 30 July 1998 and involved 12 teams consisting of five or six people each. The quality of data collection was evaluated on the basis of results from an independent check of 17 plots in the logged forests. The standard error of the standing tree volume was less than 5%.

Received 25 April; accepted 17 September 2001.

1. Nepstad, D. C. *et al.* Large-scale impoverishment of Amazonian forests by logging and fire. *Nature* **398**, 505–508 (1999).
2. Uhl, C. & Buschbacher, R. A disturbing synergism between cattle ranch burning practices and selective tree harvesting in the eastern Amazon. *Biotropica* **17**, 265–268 (1985).
3. Goldammer, J. G., Seiber, B. & Schindele, W. in *Dipterocarp Forest Ecosystems: Towards Sustainable Management* (eds Schulte, A. & Schöne, D.) 155–185 (World Scientific, Singapore, 1996).
4. Cochran, M. *et al.* Positive feedbacks in the fire dynamics of closed canopy tropical forests. *Science* **284**, 1832–1835 (1999).
5. ADB (Asian Development Bank)/BAPPENAS (National Development Planning Agency). *Causes, Extent, Impact and Costs of 1997/98 Fires and Drought*. Final Report, Annex 1 and 2. *Planning for Fire Prevention and Drought Management Project* (Asian Development Bank TA 2999-INO Fortech, Pusat Pengembangan Agribisnis, Margueles Pöyry, Jakarta, Indonesia, 1999).
6. Barber, C. V. & Schweithelm, J. *Trial by Fire 5–10* (World Resources Institute, 2000).
7. Brookfield, H., Potter, L. & Byron, Y. in *Place of the Forest: Environmental and Socio-Economic Transformation in Borneo and the Easter Malay Peninsula* 158–178 (United Nations Univ. Press, Tokyo, 1995).
8. Forestry Statistic 1998/98 (Statistik Kehutanan) 27–28 (Provincial Forestry Department of East Kalimantan (Kantor Wilayah Kehutanan dan Perkebunan Propinsi Kalimantan Timur), Samarinda, Indonesia, 1999).
9. Colfer, C., Dennis, R. A. & Applegate, G. *The Underlying Causes and Impacts of Fires in South East Asia*. Site Report 8. Long Segar, East Kalimantan Province, Indonesia, 1–88 (Center for Internal Forestry Research, Bogor, Indonesia, 2000).
10. Siegert, F. & Hoffmann, A. A. The 1998 Forest Fires in East-Kalimantan (Indonesia): A quantitative evaluation using high resolution, multitemporal ERS-2 SAR Images and NOAA-AVHRR hot spot data. *Remote Sens. Environ.* **72**, 64–77 (2000).
11. Goldammer, J. G. (ed.) *Fire in the Tropical Biot. Ecosystem Processes and Global Challenges Ecological Studies Vol. 84*, 11–28 (Springer, Berlin, 1990).
12. Kauffmann, J. B., Uhl, C. & Cummings, D. L. Fires in the Venezuelan Amazon 1: Fuel biomass and fire chemistry in the evergreen rainforest of Venezuela. *Oikos* **53**, 167–175 (1988).
13. Cochran, M. A. & Schulze, M. D. Fires as a recurrent event in tropical forests of the Eastern Amazon: Effects on forest structure, biomass, and species composition. *Biotropica* **31**, 2–16 (1999).
14. Malingreau, J. P., Stephens, G. & Fellows, L. Remote sensing of forest fires: Kalimantan and North Borneo in 1982–83. *Ambio* **14**, 314–321 (1985).
15. Leighton, M. & Wirawan, N. in *Tropical Rain Forests and the World Atmosphere* (ed. Prance, G. T.) 75–102 (AAAS Selected Symposium 101 Westview, Boulder, 1986).
16. Siegert, F. & Rücker, G. Use of multitemporal ERS-2 SAR images for identification of burned scars in south-east Asian tropical rain forest. *Int. J. Remote Sens.* **21**, 831–837 (2000).
17. Ulaby, F. T., Moore, R. K. & Fung, A. K. in *Microwave Remote Sensing: Active and Passive Vol III From Theory to Applications* 1811–1830 (Artech House, Dedham, 1986).
18. Solichin, S. H., Hinrichs, A., Soenoko, H. & Soemantri, H. *Burnt Forest Inventory Sustainable Forest Management Project*, SFMP Doc. No. 7b, 8–34 (Gesellschaft für Technische Zusammenarbeit—Sustainable Forest Management Project, Samarinda, Indonesia, 1999).
19. Holdsworth, A. R. & Uhl, C. Fire in Amazonian selectively logged rain forest and the potential for fire reduction. *Ecol. Appl.* **7**, 713–725 (1997).
20. Goldammer, J. G. Forest on fire. *Science* **284**, 1782–1783 (1999).
21. Jepson, P., Jarvie, J. K., MacKinnon, K. & Monk, K. A. The end for Indonesia's lowland rainforests? *Science* **292**, 859–861 (2001).
22. Siegert, F. & Kuntz, S. Monitoring of deforestation and land use in Indonesia with multitemporal ERS data. *Int. J. Remote Sens.* **20**(12), 2835–2853 (1999).

Acknowledgements

This research was funded by the German Federal Ministry of Economic Development and Cooperation and supported by the Gesellschaft für Technische Zusammenarbeit (GTZ) and the Indonesian Ministry of Forestry and Estate Crops, the European Space Agency (ERS European Remote Sensing Satellite 3.ERS Announcement of Opportunities), the Joint Research Centre (Space Applications Institute) and the Global Fire Monitoring Centre (Max Planck Institute for Chemistry). We gratefully acknowledge helpful comments by C. N. David and P. T. L. Ganeca for co-operating with ground survey data collection and all staff members of the participating institutions.

Correspondence and requests for materials should be addressed to F.S. (e-mail: fsiegert@zi.biologie.uni-muenchen.de).

Evolution of cooperation without reciprocity

Rick L. Riolo*, Michael D. Cohen† & Robert Axelrod‡

* Center for the Study of Complex Systems; and † School of Information; and ‡ Gerald R. Ford School of Public Policy, University of Michigan, Ann Arbor, Michigan 48109, USA

A long-standing problem in biological and social sciences is to understand the conditions required for the emergence and maintenance of cooperation in evolving populations. For many situations, kin selection¹ is an adequate explanation, although kin-recognition may still be a problem. Explanations of cooperation between non-kin include continuing interactions that provide a shadow of the future (that is, the expectation of an ongoing relationship) that can sustain reciprocity^{2–4}, possibly supported by mechanisms to bias interactions such as embedding the agents in a two-dimensional space^{4–6} or other context-preserving networks⁷. Another explanation, indirect reciprocity⁸, applies when benevolence to one agent increases the chance of receiving help from others. Here we use computer simulations to show that cooperation can arise when agents donate to others who are sufficiently similar to themselves in some arbitrary characteristic. Such a characteristic, or ‘tag’, can be a marking, display, or other observable trait. Tag-based donation can lead to the emergence of cooperation among agents who have only rudimentary ability to detect environmental signals and, unlike models of direct^{3,4} or indirect reciprocity^{9,10}, no memory of past encounters is required.

Tag-based mechanisms may be useful even if an agent is unable to observe or remember others’ actions. To show how tag-based strategies of helping can lead to the emergence of cooperation, we use the donor/recipient setting of ref. 10. In this setting agents are paired at random, and one agent has an opportunity to make a costly donation to the other. Unlike models of tag-based cooperation using an iterated ‘prisoner’s dilemma’ setting^{11–14}, this setting has no continuing interaction between pairs of agents. In Nowak and Sigmund’s model of ‘image scoring’¹⁰, an agent’s decision to donate (cooperate) is based on whether the potential recipient is known to be sufficiently generous to others. In our model, an agent’s decision to donate depends only on arbitrary ‘tags’ associated with the agents. In particular, when an agent meets another agent, it decides to help if and only if both tags are sufficiently similar. Because agents interact with only a few randomly selected others from a moderately sized population, there is little chance that a given pair will meet again. Whether or not an agent donates does not affect the likelihood of receiving help from others. Thus there is no reciprocity, either direct or indirect.

In Holland’s original formulation¹², arbitrary, evolving tags could facilitate selective interactions, and thereby be helpful for aggregation and boundary formation. While a tag-matching mechanism can be arbitrarily complex, we use the simplest mechanism. Each agent has two traits, a tag $\tau \in [0, 1]$, and a tolerance threshold $T \geq 0$. Initially, tags and tolerance levels are assigned to agents at random, uniformly sampled from $[0, 1]$. (In other experiments, we started with high, $T = 0.5$, and low, $T = 0.005$, tolerances. Except for short initial transients, the results were not substantially different from those reported here.) In each generation, each agent acts as a potential donor with P others chosen at random, with replacement. Thus for $P = 3$, each agent has three opportunities to donate, and on average is chosen three times as a potential recipient.

An agent A donates to a potential recipient B only if B’s tag is sufficiently similar to A’s tag. In particular, A donates only when B’s tag is within A’s tolerance threshold, T_A , namely when $|\tau_A - \tau_B| \leq T_A$. Thus an agent with a high T will donate to agents

with a wide range of tags, while an agent with a very small T will donate only to those whose tags are nearly the same as its own. If A does donate to B, A pays a cost, c , and B receives a benefit, b . We note that even equality of tags does not make two agents more likely to interact, but if they do interact then one will donate to the other.

After all agents have participated in all pairings in a generation, agents are reproduced on the basis of their score relative to others. The least fit, median fit, and most fit agents have respectively 0, 1 and 2 as the expected number of their offspring. This is accomplished by comparing each agent with another randomly chosen agent, and giving an offspring to the one with the higher score. (Another interpretation of this adaptive process is learning in a fixed population. In the learning interpretation, each agent compares itself to another agent, and adopts the other’s tag and tolerance if the other’s score is higher than its own.) Each offspring is subject to potential mutation which may change the offspring’s tag, tolerance or both. With probability 0.1, the offspring receives a new tag with a value drawn at random in $[0, 1]$. Also with probability 0.1, the tolerance is mutated by adding mean 0, standard deviation 0.01 gaussian noise to the old tolerance. If the new $T < 0$, it is set to 0. One run of the model consists of 100 agents and 30,000 generations. Each experimental condition is replicated 30 times.

We find that using the tag-based mechanism and adaptive processes described above, a population of agents is able rapidly to establish a substantial degree of cooperation. For example, with $P = 3$ pairings per agent per generation, and with cost $c = 0.1$ and benefit $b = 1.0$, the average donation rate was 73.6%.

Figure 1 shows the dynamics of the donation rate over the first 500 generations of a typical run using these parameters. The average payoff for the population at any time is proportional to the donation rate because each donation results in one agent gaining $b = 1.0$ and another agent losing $c = 0.1$, for a net gain to the population of 0.9. The population starts with tags and tolerances uniformly distributed, so the initial average tolerance is about 0.5, and the initial average donation rate is about 67%. Within a few generations, however, the agents with low tolerances begin to take over the population as they receive benefits from more tolerant agents, but they bear less of the cost because they donate to fewer others. By generation 70 in the run shown, the average tolerance is down to 0.020, and the donation rate is down to 43%. By chance there are some small groups of agents with similar tags and relatively low tolerances. As these agents prosper and reproduce, their off-

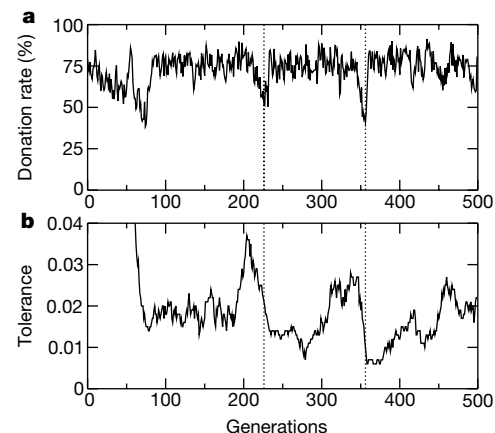


Figure 1 Population dynamics for the first 500 generations of a typical run. **a**, The donation rate. **b**, The average tolerance. Occasionally a mutant arises with a tag similar to most of the others, but with an unusually low tolerance. This mutant scores well by receiving donations from many, but donating to few. Its offspring quickly become numerous enough to lower the average donation rate and tolerance of the whole population. Soon their tag becomes the most common, resulting in a transition to a new dominant tag cluster. This happened at generation 226 and 356. We note that after these transitions, the average donation rate returned to its previous high level.

spring begin to spread through the population. Soon about 75–80% of the agents have tags that are so similar that they are within each other's tolerance range. The agents in the resulting 'dominant tag cluster' have an advantage as there are more of them to help each other. The formation of a dominant tag cluster leads to high donation rates even when averaged over the whole population. This establishes the evolution of cooperation without reciprocity.

The dynamics do not end with the establishment of a dominant cluster of similar agents who help each other. For the agents in the dominant cluster, there is only a slight selection pressure to lower their tolerance if there are no agents 'nearby', that is, with a similar tag. The average tolerance can therefore drift upwards owing to mutation occurring near the floor (base value) imposed by $T \geq 0$. Once this happens, the members of the dominant cluster are vulnerable to a relatively intolerant mutant.

The vulnerability of the dominant cluster is realized when a mutant's tag happens to be within the range of tolerance of the typical member of the dominant cluster, but the mutant's own tolerance range is small enough to prevent its donation to members of the dominant cluster. This fortunate but relatively intolerant mutant will then tend to score very well by receiving help from most of those who pair with it, while rarely giving help to others. The result is that the fortunate mutant has many offspring over the next few generations, and soon establishes a new cluster of agents with similar tags and similar (low) tolerances. As the members of this new dominant cluster do not contribute to the old cluster, the average donation rate in the population falls markedly (see Fig. 1). The members of the new cluster donate to each other when they happen to interact because, except for any further mutation of tags, they all inherit the same tag. As the new cluster grows to about 75–80% of the population, the old cluster dies out and the average donation rate rebounds. The average donation rate recovers quite quickly. This makes possible the overall donation rate of 73.6% over the entire set of 30 runs of 30,000 generations each.

We can measure the relatedness of a dominant cluster by the proportion that has its modal tag. Excluding the transient period of the first 100 generations, the relatedness of a cluster when it first becomes dominant averages 79%. Ten periods later the dominant cluster's relatedness increases to 97% as the members who give to the modal type without receiving donations from them are eliminated. Thus, by establishing dominant tag clusters, common descent has a strong influence on the maintenance of cooperation.

The new dominant cluster tends to have relatively low tolerance owing to inheritance from its founder. Over time, the average tolerance of its members tends to drift upwards. In fact, the average tolerance of a dominant cluster is much higher at its end (0.027) than its beginning (0.010). As the tolerance of a new dominant cluster increases, it becomes vulnerable to yet another relatively intolerant mutant with a similar tag. The cycle continues indefinitely. This cycle corresponds to the one found in many models of the iterated 'prisoner's dilemma' among conditional cooperators, unconditional cooperators, and defectors^{7,13–17}. In our model, the

cycle of increasing and decreasing tolerance could reflect, for example, a loss of sensory discrimination in a population when there is little selection pressure to retain it, followed by a recovery when a more discriminating individual succeeds.

The success of tag-based donation requires enough pairings per generation to establish dominant clusters. Table 1 shows the effects of varying the number of pairings per agent per generation, P , on the donation rate (the amount of cooperation) and on the population's average tolerance T . With one or two pairings, the amount of cooperation is less than 5%. With three pairings, the donation rate jumps to 74%, and then rises gradually to 79% for ten pairings. The sharp transition suggests that it may be possible to approximate these simulation patterns in an analytic model. As agents participate in more pairings, each one has a better chance of being found by an agent that will contribute to it, thus increasing the spread of agents with similar tags in future generations. Higher numbers of pairings also increases the chances that similar agents will continue to find agents to donate to, and also to receive donations, ensuring the formation of a dominant cluster with similar tags.

We note in Table 1 that when there are more than two pairings, the average level of tolerance increases, but only modestly. Thus there continues to be a pressure towards donating only to those with quite similar tags. This pressure is a result of the advantage a relatively intolerant agent has in a group of more tolerant donors with more or less similar tags. The intolerant agent gains fitness because the tolerant agents donate to it, while it bears little cost because the smaller range of tag values to which it will donate means that it will tend to donate to fewer others.

In all these simulations the typical behaviour of the system is attained within a few hundred generations, and then persists stochastically over the entire 30,000 generation period history. That full history is the basis of our reported averages.

The cost/benefit ratio, c/b , also affects the rate of donation. Table 2 shows how the donation rate and average tolerance are affected by varying c/b when P is held constant at $P = 3$, and all other parameters are unchanged. As expected, the donation rate falls when the relative cost of making donations rises. For costs less than 0.3, the rate of donation is at a high level. Beyond a cost of 0.4, the donation rate rapidly collapses.

In our model, each agent compares itself to another agent, and adopts the other's tag and tolerance if the other's score is higher than its own. Suppose instead that the agent adopts the better agent's tag and tolerance with probability proportional to how much better the other agent is. With this method, even one pairing is sufficient to achieve a donation rate of 49%, compared to only 2% shown in Table 1. The donation rate and tolerance still decrease with cost, but are less sensitive than before to increases in the number of pairings.

Our model of donation based on similarity of tags extends the insight of Nowak and Sigmund¹⁰ by reducing the requirements for the participating organisms: a potential donor incurs a certain cost in order to help another individual if and only if their tags are within the donor's range of tolerance. Tags are initially chosen at random, as are tolerances, but both are heritable and subject to mutation. Cooperation based on tag similarity does not require that the agents

Table 1 Effect of pairings on donation rate

Pairings	Donation rate (%)	Average tolerance
1	2.1	0.009
2	4.3	0.007
3	73.6	0.019
4	76.8	0.021
6	78.7	0.024
8	79.2	0.025
10	79.2	0.024

We note that increasing the number of pairings of potential donors and recipients per generation increases the donation rate. A potential donor agent A in a pair donates to a potential recipient B only if the distance between the tags of A and B is less than or equal to A's tolerance. Pairings is the number of times per generation each agent has an opportunity to donate to a randomly encountered other. The donation rate is the percentage of such encounters in which the choosing agent cooperates, that is, donates $b = 1.0$ at a cost of $c = 0.1$ to itself. We note that the donation rate increases markedly between $P = 2$ and $P = 3$, whereas the average tolerance of the population increases only slightly.

Table 2 Effect of cost of donating on donation rate

Cost	Donation rate (%)	Average tolerance
0.05	73.7	0.019
0.1	73.6	0.019
0.2	73.6	0.018
0.3	73.5	0.018
0.4	60.1	0.011
0.5	24.7	0.007
0.6	2.2	0.005

We note that increasing the cost of donating relative to the benefits conveyed decreases the donation rate. Here, the number of pairings per agent per generation is held constant at $P = 3$. When agent A donates to agent B, the recipient gets $b = 1$; the cost/benefit ratio c/b is altered by adjusting the cost, c . The donation rate is the percentage of pairings in which an agent cooperates by making a donation. Average tolerance is calculated over the entire population.

are able to recognize each other from past interactions. Nor does it require that one agent can observe and recall how the other agents behaved with third parties. Therefore cooperation on the basis of similarity could be widely applicable in situations where repeated interactions are rare, and reputations are not established. Indeed, the basis for similarity can be completely arbitrary, such as for chemical markers or cultural attributes. Cultural artefacts that can serve as tags include accents, practices or artefacts subject to fashion such as wearing hats of particular colours¹⁸. The basis for similarity also can be 'secret handshakes' or other arbitrary behavioural signals that individuals can detect¹⁹. As an agent does not have to remember previous interactions with another agent, let alone know anything about that agent's behaviour with others, an agent only needs very limited signal-detection capability. Indeed, kin recognition may use tag-based mechanisms such as the 'green beard'^{1,20–23} and 'armpit' effects^{24–28}. Using tags may also be interpreted as imposing an abstract topology on the agents in which an agent's 'neighbourhood' is defined by its tag and threshold of similarity tolerance¹⁴. In summary, our results show that cooperation can become established and be sustained even without memory. Not only do the agents not require continuing interactions, they do not even need to observe the behaviour of others or receive reports from third parties. Strategies of donating to others who have sufficiently similar heritable tags—even though such tags are initially arbitrary—can establish cooperation without reciprocity. □

Received 3 August; accepted 17 September 2001.

- Hamilton, W. D. The genetical evolution of social behaviour, I and II. *J. Theor. Biol.* **7**, 1–52 (1964).
- Trivers, R. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
- Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
- Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
- Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature* **359**, 826–829 (1992).
- Lomborg, B. Nucleus and shield: the evolution of social structure in the iterated prisoner's dilemma. *Am. Soc. Rev.* **61**, 278–307 (1996).
- Cohen, M. D., Riolo, R. L. & Axelrod, R. The role of social structure in the maintenance of cooperative regimes. *Rationality Soc.* **13**, 5–32 (2001).
- Alexander, R. D. *The Biology of Moral Systems* (Aldine de Gruyter, New York, 1987).
- Boyd, R. & Richerson, P. J. The evolution of indirect reciprocity. *Social Networks* **11**, 213–236 (1989).
- Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577 (1998).
- Axelrod, R. *The Evolution of Cooperation* 146–150 (Basic Books, New York, 1984).
- Holland, J. H. *Hidden Order: How Adaptation Builds Complexity* (Addison Wesley, Reading, Massachusetts, 1995).
- Riolo, R. L. in *Proc. 7th Int. Conf. Genetic Algorithms (ICGA97)* (ed. Bäck, T.) 378–385 (Morgan Kaufmann, San Francisco, 1997).
- Cohen, M. D., Riolo, R. L. & Axelrod, R. The emergence of social organization in the prisoner's dilemma: how context preservation and other factors promote cooperation. Working paper 99-01-002 (Santa Fe Institute, New Mexico, 1999).
- Nowak, M. A. & Sigmund, K. Oscillations in the evolution of reciprocity. *J. Theor. Biol.* **137**, 21–26 (1989).
- Lindgren, K. in *Artificial Life II* (eds Langton, C. G. et al.) 295–312 (Addison-Wesley, Reading, Massachusetts, 1991).
- Linstor, B. Evolutionary stability in the infinitely repeated prisoner's dilemma played by two-state Moore machines. *South. Econ. J.* **58**, 880–903 (1992).
- Allison, P. D. The cultural evolution of beneficent norms. *Social Forces* **71**, 279–301 (1992).
- Robson, A. J. Efficiency in evolutionary games: Darwin, Nash and the secret handshake. *J. Theor. Biol.* **144**, 379–396 (1990).
- Dawkins, R. *The Selfish Gene* 96 (Oxford Univ. Press, Oxford, 1976).
- Haig, D. Gestational drive and the green-bearded placenta. *Proc. Natl Acad. Sci. USA* **93**, 6547–6551 (1996).
- Grafen, A. Evolutionary biology—green beard as death warrant. *Nature* **394**, 521–523 (1998).
- Keller, L. & Ross, K. G. Selfish genes: a green beard in the red fire ant. *Nature* **394**, 573–575 (1998).
- Dawkins, R. *The Extended Phenotype* 146–151 (Freeman, San Francisco, 1982).
- Hauber, M. E., Sherman, P. W. & Paprika, D. Self-referent phenotype matching in a brood-parasite: the armpit effect in brown-headed cowbirds (*Molothrus ater*). *Anim. Cogn.* **3**, 113–117 (2000).
- Hauber, M. E. & Sherman, P. W. The armpit effect in hamster kin recognition. *Trends Ecol. Evol.* **15**, 349–350 (2000).
- Mateo, J. M. & Johnston, R. E. Kin recognition and the 'armpit effect': evidence of self-referent phenotype matching. *Proc. R. Soc. Lond. B* **267**, 695–700 (2000).
- Isles, A. R., Baum, M. J., Ma, D., Keverne, E. B. & Allen, N. D. Genetic imprinting—urinary odour preferences in mice. *Nature* **409**, 783–784 (2001).

Acknowledgements

For financial support we thank the Intel Corporation and the University of Michigan College of Literature, Science and the Arts Enrichment Fund. For computing facilities we thank the University of Michigan Center for the Study of Complex Systems.

Correspondence and requests for materials should be addressed to R.L.R. (e-mail: rlrlo@umich.edu).

Effects of experience and social context on prospective caching strategies by scrub jays

N. J. Emery* & N. S. Clayton†

* Sub-department of Animal Behaviour, University of Cambridge, Cambridge CB3 8AA, UK

† Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK

The authors contributed equally to this work

Social life has costs associated with competition for resources such as food¹. Food storing may reduce this competition as the food can be collected quickly and hidden elsewhere^{2–4}; however, it is a risky strategy because caches can be pilfered by others^{5–9}. Scrub jays (*Aphelocoma coerulescens*) remember 'what', 'where' and 'when' they cached^{10–13}. Like other corvids^{6–9,14}, they remember where conspecifics have cached, pilfering them when given the opportunity, but may also adjust their own caching strategies to minimize potential pilfering. To test this, jays were allowed to cache either in private (when the other bird's view was obscured) or while a conspecific was watching, and then recover their caches in private. Here we show that jays with prior experience of pilfering another bird's caches subsequently re-cached food in new cache sites during recovery trials, but only when they had been observed caching. Jays without pilfering experience did not, even though they had observed other jays caching. Our results suggest that jays relate information about their previous experience as a pilferer to the possibility of future stealing by another bird, and modify their caching strategy accordingly.

In the wild, food-storing corvids return to caches that they had hidden in the presence of conspecifics, and readily re-cache them in new places when the observers are no longer present (for example ravens^{8,9}, European jays¹⁵, scrub jays; N.S.C., unpublished observations). We proposed that birds re-cache to minimize potential pilfering by observers. We therefore predicted that they would be more likely to re-cache any uneaten food, and specifically in new sites unbeknown to an observer, but only if they had been watched during the caching trial. To test this hypothesis, scrub jays were allowed to cache wax worms in a sand-filled caching tray during two

Table 1 Behaviour of the observer + pilferer birds during observed and in private caching treatments

Behaviour	Caching treatment		Wilcoxon pairs test		
	Observed	In private	n	Z	P
No. cached					
Davis	8.19 ± 1.55	4.71 ± 0.81	7	2.37	<0.05
Cambridge	10.48 ± 3.43	9.10 ± 3.05	7	0.51	>0.5
No. recovered					
Davis	4.61 ± 0.93	3.95 ± 0.84	7	0.85	>0.1
Cambridge	5.38 ± 1.56	4.19 ± 1.10	7	0.08	>0.5
Proportion recovered					
Davis	0.71 ± 0.06	0.57 ± 0.07	7	2.03	<0.05
Cambridge	0.56 ± 0.08	0.70 ± 0.09	7	1.69	>0.05
Recovery accuracy*					
Davis	2.21 ± 0.46	3.21 ± 1.01	7	1.15	>0.1
Cambridge	3.07 ± 0.92	1.52 ± 0.24	7	1.36	>0.1
No. re-cached					
Davis	2.19 ± 0.68	0.57 ± 0.32	7	2.20	<0.05
Cambridge	2.74 ± 1.01	0.36 ± 0.19	7	2.20	<0.05
Proportion re-cached					
Davis	0.44 ± 0.20	0.06 ± 0.03	7	2.20	<0.05
Cambridge	0.28 ± 0.07	0.08 ± 0.04	7	2.20	<0.05

Treatments consisted of three trials/caching treatment at Davis, followed by three trials/caching treatment at Cambridge (data are mean ± s.e.m.). A Wilcoxon matched-pairs test compared the effect of caching treatment for each of the behaviours listed.

* Number of looks to find first cache.

experimental conditions: while being observed by a conspecific (observed), and while the observer's view was blocked (in private). The storers recovered their caches 3 h later. During these recovery trials, which were always conducted out of sight of the observer, the birds could eat any caches that they recovered, but they could also re-cache them. They were given access to a new caching tray as well as the original one, so that they could re-cache in both new and old sites. New trays were used on each caching trial so that the cache sites were unique for each trial.

Scrub jays in the observer + pilferer group had participated in an earlier observational learning experiment in which they observed a conspecific cache, and subsequently pilfered those caches¹⁴. These birds received a series of caching and recovery trials at the University of California, Davis. The design was replicated at the University of Cambridge, UK, using the same birds.

Table 1 summarizes the main findings and statistics. Although birds cached significantly more items when observed by a conspecific at Davis, this effect was not found in the replicate at Cambridge, nor in subsequent tests (see below). The pattern of results for all other behaviours is very similar for both Davis and Cambridge. There was no effect of caching treatment on the number of items

recovered, even when we analysed the number of items recovered as a proportion of the number cached. There was also no difference in the birds' accuracy at cache recovery, as measured by the number of searches birds made to find the first cache.

The principal result is that birds re-cached significantly more items during recovery when they had been observed during caching. As the number of items that can be re-cached depends on the number originally cached and subsequently recovered, we calculated the number of worms that were re-cached as a proportion of the number of worms that were recovered per bird. The difference between the observed and in private caching treatments was also found for the proportion of worms re-cached, a measure that controls for differences in the number of items cached and recovered. These results therefore confirmed our prediction that birds would re-cache significantly more if they had been observed caching. We also predicted that caching treatment would influence a bird's choice of re-caching location. Figure 1a, b (Davis and Cambridge, respectively) supports this prediction by showing that re-caching was predominantly in new sites rather than old sites when the bird had been observed during caching. By contrast, birds that had been allowed to cache in private did not discriminate between old and new sites.

Birds in the observer + pilferer group had to remember whether they had been watched during caching to know when to re-cache during recovery, and whether to re-cache in new sites. Being observed recently during caching could lead to a generalized tendency to re-cache when presented with the opportunity to do so at a later time. Alternatively, the bird may remember whether it had been watched while caching in a specific tray. We discriminated between these two possibilities by testing whether the observer + pilferer group could keep track of the social context of previous caching episodes that occurred in close temporal proximity. We used interleaved observed and in private caching trials: the birds cached in one tray while being observed by a conspecific (observed tray), and then cached in another tray in private immediately afterwards (in private tray), and vice versa. After a 3-h retention interval, the birds received a recovery trial in which they were presented simultaneously with both the observed tray and the in private tray in which they had cached, as well as a new tray.

Table 2 shows the mean ($\pm$ s.e.m.) number of items cached, recovered and re-cached, the number of looks to find the first cache and the proportion of caches recovered and re-cached per tray (observed or in private) during the interleaved trials. Note that the numbers of items cached are much lower than those shown in Table 1, presumably as a consequence of using interleaved trials in which the second caching episode occurred immediately after the first one. In line with the previous results for the observer + pilferer group in Cambridge, however, we found no difference in the number of items cached in the observed tray and in the private tray. The number of items recovered from each tray is also less than that shown in Table 1, probably because the birds had a choice of two trays from which they could recover as opposed to only one, yet the duration of the recovery trial was the same. Birds recovered

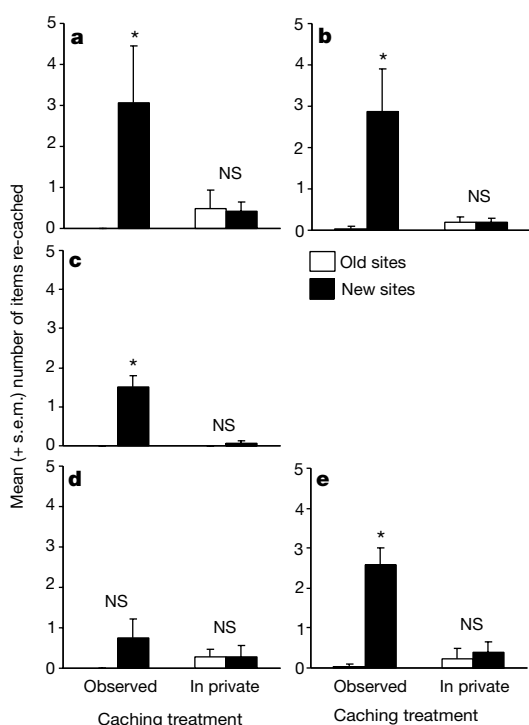


Figure 1 Mean ($\pm$ s.e.m.) number of worms re-cached by the three groups. **a**, The observer + pilferer group in Davis re-cached significantly more in the new sites during the observed caching treatment ($n = 7$, $Z = 2.20$, $P < 0.05$), but failed to discriminate between new and old sites during the in private caching treatment ($n = 7$, $Z = 0$, $P = 1.0$). **b**, The observer + pilferer group in Cambridge also re-cached significantly more in new sites when observed during caching ($n = 7$, $Z = 2.20$, $P < 0.05$), but again did not discriminate between re-caching in old or new sites during the in private caching treatment (no statistical comparison possible). **c**, The observer + pilferer group also selectively re-cached food in new sites in the observed tray ($n = 7$, $Z = 2.37$, $P < 0.05$), but did not discriminate between old and new sites in the in private tray during the interleaved caching trials (no statistical comparison possible). **d**, The observer group did not selectively re-cache in either new or old sites during both the observed and the in private caching treatments (old: $n = 7$, $Z = 1.6$, $P > 0.1$; new: no statistical comparison possible). **e**, The pilferer group re-cached significantly more food items in new sites during the observed caching treatment ($n = 7$, $Z = 2.37$, $P < 0.05$), but as with the observer + pilferer group, did not re-cache selectively in either old or new sites during the in private caching treatment (no statistical comparison possible). All analyses were Wilcoxon matched-pairs tests. Asterisk, $P < 0.05$, NS, not significant ($P > 0.05$).

Table 2 Behaviour of the observer + pilferer birds in the observed tray and in private tray during the interleaved trials

Behaviour	Caching treatment		Wilcoxon pairs test		
	Observed	In private	<i>n</i>	<i>Z</i>	<i>P</i>
No. cached	3.93 $\pm$ 0.77	3.57 $\pm$ 0.76	7	0.25	>0.5
No. recovered	3.00 $\pm$ 0.53	1.21 $\pm$ 0.32	7	1.99	<0.05
Proportion recovered	0.73 $\pm$ 0.07	0.39 $\pm$ 0.09	7	2.37	<0.05
Recovery accuracy*	4.21 $\pm$ 2.03	4.57 $\pm$ 0.85	7	0.42	>0.5
No. re-cached	2.00 $\pm$ 0.68	0.14 $\pm$ 0.14	7	2.20	<0.05
Proportion re-cached	0.45 $\pm$ 0.11	0.04 $\pm$ 0.04	7	2.37	<0.05

A Wilcoxon matched-pairs test compared the effect of caching treatment for each of the behaviours listed (data are mean $\pm$ s.e.m.).

* Number of looks to find first cache.

significantly more items from the observed tray. We found also that the birds recovered proportionally more items from the observed tray. There were no differences in recovery accuracy, however.

Importantly, Table 2 also shows that birds re-cached more of those items that were recovered from the observed tray than from the in private tray, even when the data were analysed as a proportion of the number of items recovered (which adjusts for differences in recovery). Figure 1c shows that re-caching was predominantly made in new sites in the observed tray. By contrast, the birds showed little tendency to discriminate between re-caching in old and new sites in the in private tray. This pattern of results is consistent with those obtained from the previous study. It suggests that jays remember not only whether they were being observed, but can also recall the specific tray in which they cached while being watched, rather than relying on a default mechanism to re-cache if they have been observed recently.

Although some food-storing species avoid caching in those areas from which food has been stolen if the caches are pilfered regularly^{16,17}, this is not observed after a single pilfering event¹⁸, which suggests that birds may need to learn about the costs and benefits of pilfering. We therefore assessed the role of prior learning by testing the cache recovery behaviour of the observer group that had experience of observing a conspecific cache, but no experience of subsequently pilfering those caches. Birds in the observer group had the same amount of caching and recovery experience as the observer + pilferer birds, and considerable experience of observing the caching behaviour of conspecifics. The observer group had previously observed birds caching during the observational learning study¹⁴ and they served also as observers in the studies described earlier. However, they differed from the observer + pilferer group in that they never had the opportunity to pilfer another bird's caches.

Table 3 displays the mean ($\pm$ s.e.m.) number of worms cached, recovered and re-cached, the number of looks to find the first cache and the proportion of caches recovered and re-cached during the observed or in private caching treatments for the observer group. It is clear from Table 3 that there is no difference in the number of items cached or recovered between the observed and in private caching treatments, and an analysis of the proportion recovered also failed to find any significant difference. There was also no effect of caching treatment on cache accuracy. Unlike the observer + pilferer group, however, the observer group did not display any significant difference in the number of items re-cached between the observed and in private caching treatments or the proportion of items re-cached.

Figure 1d shows that, in contrast to the observer + pilferer group,

Table 3 Behaviour of observer and pilferer birds during observed and in private caching treatments

Behaviour	Caching treatment		Wilcoxon pairs test		
	Observed	In private	<i>n</i>	<i>Z</i>	<i>P</i>
No. cached					
Observer group	6.38 $\pm$ 0.80	4.62 $\pm$ 0.43	7	1.57	>0.1
Pilferer group	5.53 $\pm$ 1.50	7.05 $\pm$ 1.95	7	1.01	>0.1
No. recovered					
Observer group	3.36 $\pm$ 0.53	2.69 $\pm$ 0.30	7	1.10	>0.1
Pilferer group	3.67 $\pm$ 0.78	4.00 $\pm$ 0.97	7	1.35	>0.1
Proportion recovered					
Observer group	0.59 $\pm$ 0.08	0.58 $\pm$ 0.07	7	0.17	>0.5
Pilferer group	0.73 $\pm$ 0.08	0.62 $\pm$ 0.06	7	1.35	>0.1
Recovery accuracy*					
Observer group	2.29 $\pm$ 0.53	2.31 $\pm$ 0.41	7	0.34	>0.5
Pilferer group	2.62 $\pm$ 0.78	2.12 $\pm$ 0.53	7	0.73	>0.1
No. re-cached					
Observer group	0.69 $\pm$ 0.43	0.29 $\pm$ 0.21	7	1.60	>0.1
Pilferer group	2.09 $\pm$ 0.50	0.62 $\pm$ 0.42	7	2.20	<0.05
Proportion re-cached					
Observer group	0.10 $\pm$ 0.04	0.06 $\pm$ 0.04	7	1.60	>0.1
Pilferer group	0.44 $\pm$ 0.09	0.09 $\pm$ 0.06	7	2.37	<0.05

A Wilcoxon matched-pairs test compared the effect of caching treatment for each of the behaviours listed (data are mean $\pm$ s.e.m.).

* Number of looks to find first cache.

birds in the observer group re-cached very little in both the observed and in private caching treatments. They did not selectively re-cache in new sites, and made too few re-caches to permit a valid statistical analysis (Table 3). This pattern of results suggests that birds need pilfering experience to know when to re-cache because the critical difference between the observer and observer + pilferer groups is whether they had the opportunity to pilfer another bird's caches.

To verify this hypothesis, and to test whether experience of observing caching behaviour is also critical for prospective caching, a further group of birds was tested. Birds in the pilferer group had the opportunity to listen to, but not observe, another bird caching, and were then allowed to pilfer those caches in the earlier observational learning study¹⁴. As all the birds had received extensive caching experience, they readily searched for caches even when they had not seen the storer hide food^{10–13}. These birds differed therefore from observer and observer + pilferer birds in not having the opportunity to observe conspecifics caching in these experiments. In common with the observer + pilferer group, but in contrast to the observer group, the pilferer group had previous experience of pilfering. If pilfering, but not observation, is the critical experience for knowing when to re-cache, then birds of the pilferer group should show the same pattern of results as the observer + pilferer group.

Table 3 also displays the mean ($\pm$ s.e.m.) number of worms cached, recovered and re-cached, the number of looks to find the first cache and the proportion of caches recovered and re-cached during either observed or in private caching treatments for the pilferer group. We found no significant difference between the two caching treatments in the number of items cached or recovered. There was also no significant difference between the observed and in private caching treatments in the proportion of items recovered or accuracy to find a cache. However, in line with the observer + pilferer group, and in contrast to the observer group, the number of items re-cached was significantly greater in the observed caching treatment. An analysis of the proportion of items re-cached confirmed this result.

Figure 1e shows that the pilferer group re-cached more items during recovery if they had been observed during caching, and predominantly in new sites. There was no preference for caching in new sites when they had previously cached in private. Taken together, this pattern of results demonstrates that birds need experience of being a pilferer to know when to re-cache, and whether to re-cache in new sites, but they do not need experience of being an observer.

To our knowledge, this is the first experimental demonstration that a non-human animal can remember the social context of specific past events, and adjust their present behaviour to avoid potentially detrimental consequences in the future, in this case pilfering. To do this, scrub jays need experience of pilfering another bird's caches, but do not require experience of observing a conspecific hide food. They can recall specific past events^{10–13}, but the present results raise the possibility that they can also plan for the future. The jays seem to have transferred their previous experience of being a pilferer to the current situation in which their own caches might be stolen. This may be a good candidate for knowledge attribution in conspecifics¹⁹ (seeing leads to knowing), use of this knowledge to influence subsequent behaviour²⁰ (re-caching in new locations) or even tactical deception²¹. Mental time travel (episodic memory and future planning) and mental attribution were thought to be unique to humans^{22,23}. The cache recovery model presents a new way of addressing these issues in animals. □

Methods

Subjects

A total of 21 adult hand-raised scrub jays were used, all of whom had been tested in previous studies of caching behaviour²⁴, memory^{10–13} and observational learning¹⁴. All of the birds had the same experience of caching and recovering food. The only difference

between the birds was their experience in the earlier observational learning study¹⁴—observer, pilferer or observer + pilferer. These categories are directly related to the specific experiences in the previous experiment, not the developmental histories of the birds. Although we cannot state that birds in the pilferer group had never observed other birds cache they did not do so in the previous study or in the current one. Importantly, the observer birds have never pilfered another bird's caches.

The observer + pilferer ($n = 7$) group was first tested at the University of California, Davis (19–26 July 2000), with a replicate study performed at the University of Cambridge (15 January to 16 February 2001). Birds in the observer ($n = 7$) and pilferer ($n = 7$) groups were tested at Cambridge (15 January to 16 February 2001). The interleaved trials study (observer + pilferer group only) was performed at Cambridge (27–28 February 2001). All of the birds were housed individually in cages ($45 \times 76 \times 76$ cm). In Davis, birds were housed in an outdoor aviary. In addition to natural lighting, we provided fluorescent strip lighting. In Cambridge, the birds were housed in the same cages as Davis, but these were placed in a quarantined indoor room in which only fluorescent lighting was provided. Birds were fed a mixture of Iams mini-chunk dog food biscuits and peanuts, both of which were provided in powdered form to ensure that the birds could not cache outside of the experiment. Water was provided *ad libitum*.

Apparatus

The fronts of the individual cages were made of aluminium wire and the sides were made of solid aluminium. The storer was placed in a cage located adjacent to a second cage containing the observer, with a 20-cm-wide gap between the two cages. On the back of each cage was a 25 cm² perspex panel, which allowed the storer and observer to see one another. Caching trays^{10–13} were constructed from sand-filled plastic ice-cube trays, each containing a 2×8 array of moulds that were attached to a wooden board. Each caching tray was made unique by attaching different configurations of Lego (Netfield) bricks onto the board behind the ice-cube tray.

Procedure

Birds were deprived of food overnight. Caching trials started at 10:00 (Davis) or 11:00 (Cambridge) the following day. During every 15-min caching trial ($n = 3$ per caching treatment, observer + pilferer group in Davis and observer, pilferer and observer + pilferer groups in Cambridge; $n = 2$ pairs, observer + pilferer group interleaved trials in Cambridge) each subject received one sand-filled caching tray and a bowl containing 50 wax worms. During the unobserved caching treatment (in private), a towel covered the back of the cage so that the view of the observer bird was completely obscured. During the observed caching treatment, the observer had a clear view of the storer. At the end of each caching trial, the tray and food bowl were removed from each cage. The experimenters recorded the number and location of wax worms cached in the tray. Any extraneous food that the birds did not cache was removed. The number of worms cached therefore refers to the number of worms that remained in the tray at the end of the caching trial. Observers were given 15 min in which they were allowed to eat their maintenance diet at the end of the caching trial, and each bird was given three wax worms. In Davis and Cambridge, two caching trays were placed inside the subject's cage during recovery trials. One of the trays was unfamiliar to the bird. The other tray contained the previously cached worms and was placed in its original location. Three caching trays were placed inside the subject's cage during recovery trials for the interleaved trials study. One tray was unfamiliar to the bird, whereas the other two trays contained the previously cached worms, which were placed in the same locations as they were placed during the observed tray and in private tray caching treatments.

The subjects were allowed to recover the cached food items for 10 min. The number and location of the searches was recorded by direct observation, as well as the number of caches recovered and whether these were eaten or re-cached. We also noted the location of the re-cached food items, and whether they were made in new sites compared to those places the bird had cached in during the previous caching trial (old site). We calculated the number of searches to find the first cache (recovery accuracy), the proportion of worms recovered and the proportion of those recoveries that were re-cached. All 21 birds had previous experience with the presence of towels on their cages for at least three trials each. They were restricted from caching during these trials, but were given powdered food and three wax worms.

Analysis

We used the Wilcoxon matched-pairs test on the mean values summed across trials for each caching treatment. As the number of items recovered depends on the number of items cached, and the number of items re-cached depends on the number of caches that are recovered, we also analysed the proportion of caches that were recovered and the proportion of those re-cached. For all analyses, alpha was set at 0.05.

Received 23 July; accepted 20 September 2001.

- Zentall, T. R. & Galef, B. G. (eds) *Social Learning. Psychological and Biological Perspectives* (Lawrence Erlbaum Associates, London, 1998).
- Shettleworth, S. J. in *Behavioral Brain Research in Naturalistic and Semi-naturalistic Settings* (eds Alleva, E., Fasolo, A., Lipp, H.-P. & Nadel, L.) 158–179 (Kluwer Academic, The Hague, 1995).
- Vander Wall, S. B. *Food Hoarding in Animals* (Univ. Chicago Press, Chicago, 1990).
- Clarkson, K., Eden, S. F., Sutherland, W. J. & Houston, A. I. Density dependence and magpie food hoarding. *J. Anim. Ecol.* **55**, 111–121 (1986).
- Gibb, J. A. Populations of tits and goldcrests and their food supply in pine populations. *Ibis* **102**, 163–208 (1960).
- Bednekoff, P. A. & Balda, R. P. Observational spatial memory in Clark's nutcrackers and Mexican jays. *Anim. Behav.* **52**, 833–839 (1996).

- Heinrich, B. & Pepper, J. W. Influence of competitors on caching behavior in the common raven, *Corvus corax*. *Anim. Behav.* **56**, 1083–1090 (1998).
- Bugnyar, T. & Kotrschal, K. Do ravens manipulate the others' attention in order to prevent or achieve social learning opportunities? *Adv. Ethol.* **36**, 106 (2001).
- Heinrich, B. *Mind of the Raven* (Harper Collins, New York, 1999).
- Clayton, N. S. & Dickinson, A. D. Episodic-like memory during cache recovery by scrub jays. *Nature* **395**, 272–278 (1998).
- Clayton, N. S. & Dickinson, A. D. Scrub jays (*Aphelocoma coerulescens*) remember the relative time of caching as well as the location and content of their caches. *J. Comp. Psychol.* **113**, 403–416 (1999).
- Clayton, N. S. & Dickinson, A. D. Memory for the contents of caches by Scrub Jays. *J. Exp. Psychol. Anim. Behav. Proc.* **25**, 82–91 (1999).
- Clayton, N. S., Yu, K. & Dickinson, A. D. Scrub jays (*Aphelocoma coerulescens*) can form integrated memory for multiple features of caching episodes. *J. Exp. Psychol. Anim. Behav. Proc.* **27**, 17–29 (2001).
- Clayton, N. S., Griffiths, D. P., Emery, N. J. & Dickinson, A. D. Episodic-like memory in animals. *Phil. Trans. R. Soc. Lond. B* **356**, 1483–1491 (2001).
- Goodwin, D. Further observations on the behaviour of the jay. *Ibis* **98**, 186–219 (1956).
- Stevens, T. A. & Krebs, J. R. Retrieval of stored seeds by marsh tits (*Parus palustris*) in the field. *Ibis* **128**, 513–515 (1984).
- Hampton, R. R. & Sherry, D. F. The effects of cache loss on choice of cache sites in the black-capped chickadee. *Behav. Ecol.* **5**, 44–50 (1994).
- Baker, M. C. & Anderson, P. Once-pilfered cache sites not avoided by black-capped chickadees. *Anim. Behav.* **49**, 1599–1602 (1995).
- Hare, B., Call, J., Agnetta, B. & Tomasello, M. Chimpanzees know what conspecifics do and do not see. *Anim. Behav.* **59**, 771–785 (2000).
- Hare, B., Call, J. & Tomasello, M. Do chimpanzees know what conspecifics know? *Anim. Behav.* **61**, 139–151 (2001).
- Whiten, A. & Byrne, R. W. Tactical deception in primates. *Behav. Brain Sci.* **11**, 233–244 (1988).
- Suddendorf, T. & Corballis, M. C. Mental time travel and the evolution of the human mind. *Genet. Soc. Gen. Psychol. Monogr.* **123**, 133–167 (1997).
- Heyes, C. M. Theory of mind in nonhuman primates. *Behav. Brain Sci.* **21**, 101–148 (1998).
- Clayton, N. S. & Dickinson, A. D. Motivational control of caching behaviour in the scrub jay, *Aphelocoma coerulescens*. *Anim. Behav.* **57**, 435–444 (1999).

Acknowledgements

This work was supported by the National Institutes of Health, National Institute on Aging and Whitehall foundation grants to N.S.C. N.J.E. was funded by a Medical Research Council programme grant to E. B. Keverne. We thank S. Hettige, D. Jennings and V. R. Metcalf for help in running the experiments. We also thank the University of California, Davis, for allowing us to conduct the first trials there and for providing the necessary facilities. We thank S. Baron-Cohen, E. B. Keverne, K. N. Laland, D. I. Perrett and C. J. Saldanha for comments on the manuscript. We are especially grateful to T. J. Bussey, A. Dickinson and N. J. Mackintosh for discussion and comments on the manuscript.

Correspondence and requests for materials should be addressed to N.S.C. (e-mail: nsc22@cam.ac.uk).

The central nervous system stabilizes unstable dynamics by learning optimal impedance

Etienne Burdet*†‡§, Rieko Osu*†‡, David W. Franklin§¶, Theodore E. Milner† & Mitsuo Kawato*‡§

* Department of Mechanical Engineering, National University of Singapore, 119260, Singapore

† Kawato Dynamic Brain Project, ERATO, JST, Hilaridai, Seika-cho, Soraku-gun, Kyoto, 619 0288, Japan

§ ATR Human Information Science Laboratories, Hilaridai, Seika-cho, Soraku-gun, Kyoto, 619 0288, Japan

¶ School of Kinesiology, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

‡ These authors contributed equally to the work

To manipulate objects or to use tools we must compensate for any forces arising from interaction with the physical environment. Recent studies indicate that this compensation is achieved by learning an internal model of the dynamics^{1–6}, that is, a neural representation of the relation between motor command and

movement^{5,7}. In these studies interaction with the physical environment was stable, but many common tasks are intrinsically unstable^{8,9}. For example, keeping a screwdriver in the slot of a screw is unstable because excessive force parallel to the slot can cause the screwdriver to slip and because misdirected force can cause loss of contact between the screwdriver and the screw. Stability may be dependent on the control of mechanical impedance in the human arm because mechanical impedance can generate forces which resist destabilizing motion. Here we examined arm movements in an unstable dynamic environment created by a robotic interface. Our results show that humans learn to stabilize unstable dynamics using the skilful and energy-efficient strategy of selective control of impedance geometry.

How do we succeed in performing mechanically unstable tasks? One way would be to stabilize the arm by means of feedback control. However, neural feedback, whether operating through involuntary (reflex) or voluntary commands to our muscles, is delayed by the sensory feedback pathways, which would tend to increase rather than reduce instability^{10,11}. Alternatively, stability might be achieved by controlling mechanical impedance (resistance to imposed motion); specifically, the spring-like property of muscles^{12,13}. Demonstrating impedance control is not trivial, because it is necessary to show that the shape or orientation of the impedance can be voluntarily modified independently of the force applied by the hand. Modifications in impedance that depend on changes in applied force do not constitute proof of impedance control because muscle stiffness inherently scales with activation level^{14,15}. To demonstrate impedance control, we compared multi-joint arm movements in a null field (NF) and in a divergent force field (DF), which produced an unstable interaction with the arm, but required no change in applied force relative to the NF.

Subjects made horizontal point-to-point movements away from the body, along the y axis of our coordinate system (Fig. 1a). The hand was linked by means of a stiff brace to a handle at the end of a robotic manipulandum (PFM) that exerted computer-controlled forces during movement (see Methods). The DF produced a negative elastic force perpendicular to the target direction. The

robot produced no force when trajectories followed the y axis, but the hand was pushed away whenever it deviated from the y axis (red arrows in Fig. 1a). The force (F_x, F_y) exerted on the hand was

$$\begin{bmatrix} F_x \\ F_y \end{bmatrix} = \begin{bmatrix} \beta x \\ 0 \end{bmatrix} \quad (1)$$

where $\beta > 0$ (N m^{-1}) was chosen to be larger than the stiffness of the arm measured in NF movements so as to produce instability.

The NF movements (green in Fig. 1b) had trajectories that were approximately straight. Almost all trials ended on target and the distribution of endpoints was bell-shaped. However, the initial movement direction varied slightly from trial to trial owing to motor output variability. The DF amplifies such variations by pushing the hand with a force proportional to the deviation from the y axis (red arrows in Fig. 1a). Consequently, the initial trials in the DF exhibited unstable behaviour, diverging widely either to the right or to the left. Most trials ended either on the left or right safety barrier (black vertical lines in the left plot of Fig. 1c) with a roughly symmetrical distribution of positions. With practice, however, subjects gradually became proficient at producing straighter trajectories, similar to those in the NF (compare Fig. 1b and the middle plot of Fig. 1c). Most trials then ended on target with a bell-shaped endpoint distribution. The hand-path error (see Methods, equation (2)) decreased significantly between the first five trials and the last five trials ($P = 0.002$; Fig. 2a), indicating that the subjects learned to adapt to the dynamics of the DF. At the end of the learning trials, no significant differences in the hand-path error were found between

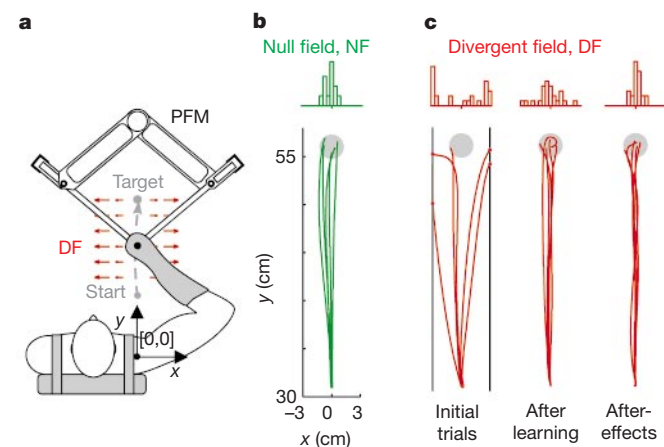


Figure 1 Experimental set-up and learning in a divergent force field (DF). **a**, The PFM exerts forces on the hand during horizontal point-to-point arm movements away from the body. The red arrows graphically depict the DF. **b**, Five consecutive trajectories of one subject are shown for movements in a null force field (NF, green). At the top of trajectory plot, a histogram shows the distribution of endpoint positions (based on curvature greater than 0.03 m^{-1}) for the corresponding five movements of all subjects. **c**, Learning in the DF (red). Five trajectories of the same subject are shown, which include the first five trials (left), the last five trials (middle), and five consecutive after-effect trials (right). For safety reasons the DF was turned off when the trajectory deviated more than 3 cm from the y axis. The black lines indicate this safety zone.

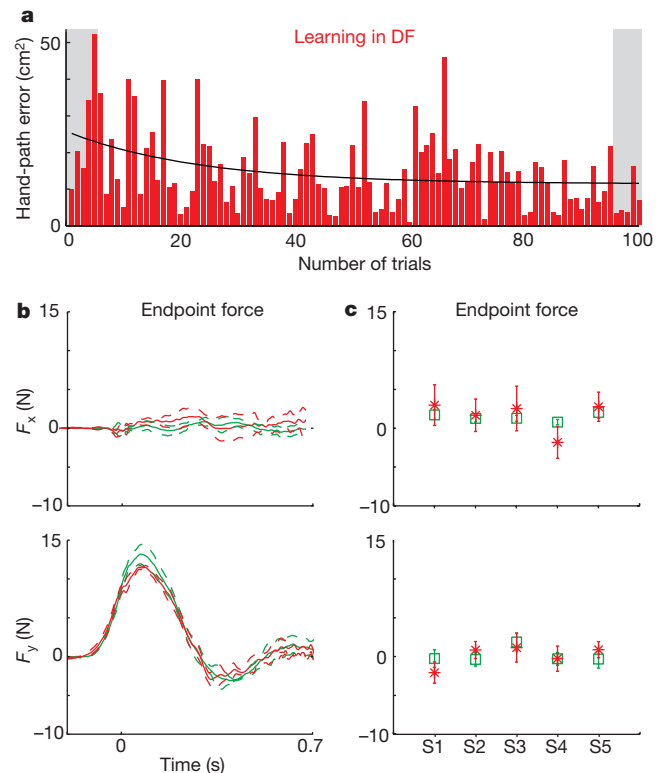


Figure 2 Error evolution during learning the DF, endpoint force–time profiles after learning for the same subject as in Fig. 1, and forces at the movement midpoint for all subjects. The hand-path error is defined in the Methods. **a**, Hand-path error as a function of the number of learning trials for the DF. The solid curve is a least-squares fit of an exponential function to the data. **b**, Mean endpoint force profiles with standard deviations in the x and y directions after learning, for NF (green) and DF (red). The force is similar in the DF and NF. **c**, The mean x and y forces and standard deviations at the midpoint of the trajectory in the NF (green) and DF (red) for all five subjects.

movements in the NF and DF ($P = 0.906$), indicating that movements had become stable and were similar under the two conditions. This is also indicated by the force applied to the PFM by the arm in the DF, which was not significantly different from that in the NF, after learning (Fig. 2b, c).

To examine the after-effects of learning the DF, the force field was unexpectedly removed on selected trials after learning. These after-effect movements were characterized by trajectories that hardly deviated from the y axis with endpoints concentrated in the middle of the target (Fig. 1c, right). The hand-path error revealed that they were even straighter than NF trajectories ($P = 0.001$).

How did subjects overcome the instability of their movements? Because of the unpredictability of the direction of the disturbance a forward or inverse dynamics model^{5,7} could not have been used. In contrast, an increase in mechanical impedance would provide increased resistance to a disturbance regardless of its direction. Perhaps subjects used generalized muscle coactivation to achieve an isotropic increase in stiffness at the hand. This has been demonstrated previously during posture¹⁶. An analogous strategy is commonly employed in single-joint tasks when the environment is destabilizing^{17,18}. On the other hand, the stiffness might have been selectively controlled in the direction of the instability. Such control of limb impedance was first proposed 16 years ago¹², but has yet to be demonstrated.

These possibilities were investigated by examining the stiffness at the hand after learning (see Methods). We measured stiffness around the midpoint of movement paths, using a technique¹⁹ that provides an unbiased and accurate estimate of stiffness and requires fewer trials than previous methods^{20,21}. On random trials, the PFM briefly displaced the hand by a constant distance from a prediction of the undisturbed trajectory¹⁹. Stiffness was estimated from the restoring force, produced in response to the displacement, divided by the amplitude of the displacement.

Stiffness at the hand in the DF was modified in shape and orientation relative to stiffness in the NF (Fig. 3, compare red and green data). Although the x component of stiffness was significantly larger in the DF than in the NF (Fig. 4a, compare red and green data), the y component of stiffness in the DF was not significantly different from that in the NF (Fig. 4b, compare red and green data).

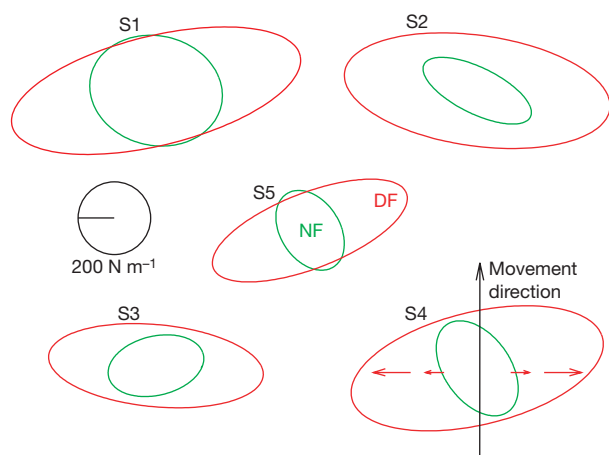


Figure 3 Adaptation of stiffness geometry to the DF. Stiffness ellipses—the graphical depiction of the elastic restoring force corresponding to a unit displacement of the hand—are shown for five subjects in the NF and DF. The long and short axes of the ellipse represent the directions of maximal and minimal stiffness, respectively. The stiffness in the DF (red) changed relative to the NF (green). We note that in the DF the stiffness increased significantly in the direction of the instability (shown by small red arrows for subject S4), but there was relatively little change along the movement direction (see also Fig. 4).

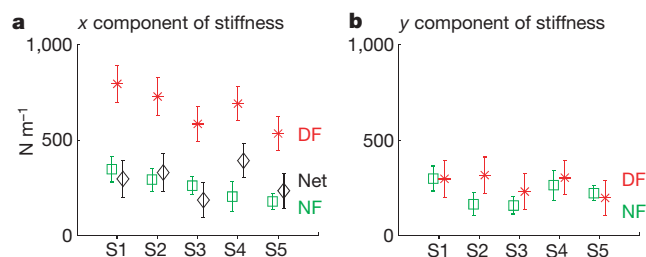


Figure 4 Stiffness and stability. **a**, The x components of hand stiffness in the NF (green), hand stiffness in the DF (red) and net stiffness of the hand and DF environment (Net, black) after learning. Vertical bars show 90% confidence intervals. **b**, The y components of hand stiffness in the NF (green) and in the DF (red).

There was a selective increase in stiffness in the direction of instability. After learning, the stiffness of the arm increased sufficiently that the combined stiffness of the arm and the environment in the x direction (Fig. 4a, Net, black data) was positive and similar in magnitude to that measured during NF movements (Fig. 4a, NF, green data). Trial-to-trial variations in x force did not correlate with changes in stiffness in the x direction for any subject (Table 1), providing even stronger support for impedance control. Furthermore, we can conclude that the adaptation was preprogrammed and not reactive to the unstable environment because the DF was not applied on those trials used for stiffness measurement. (This, however, was not detected by the subjects.)

Previous impedance adaptation studies did not address the capacity to modify stiffness geometry independently of force. Adaptive increases in stiffness observed for single-joint tasks in unstable environments^{17,18,22} provide a limited view of the adaptive capabilities of the central nervous system (CNS) because the only parameter that can be varied is stiffness magnitude. Changes in stiffness geometry, observed in ball catching²⁰, occurred principally while joint torque was changing to compensate for the momentum of the ball. The anticipatory increase in stiffness just before ball impact involved generalized muscle coactivation, which serves mainly to increase stiffness magnitude. The use of biarticular shoulder muscles to perform a postural task, requiring no net shoulder torque⁹, might be considered a form of impedance control. However, because joint stiffness clearly scaled with joint torque in this task it is not clear whether the involvement of biarticular muscles represents genuine impedance control or ‘hard-wired’ muscle synergy. Similarly, the statistically significant covariation of change in stiffness with change in external force found for movements in a velocity-dependent force field²³ could be attributed largely to the linear relationship between joint torque and stiffness^{14,15}. The modifications of stiffness observed during movements constrained by a mechanical channel²⁴ were also probably due to changes in force. In contrast, adaptation to the DF required no change in force or joint torque relative to NF movements, yet subjects skilfully adapted their stiffness to the instability of the environment so that stiffness increased only in the required direction and by the required amount. Using the DF, we have been able to show for the first time that the CNS can voluntarily control the magnitude, shape and orientation of the endpoint stiffness in a predictive way that is independent of the force needed to compensate for the imposed dynamics.

Table 1 Correlation between stiffness and force

Subject	S1	S2	S3	S4	S5
R^2	0.24	0.53	0.08	0.29	0.34

R^2 is the correlation between the change in stiffness and the change in force in the DF relative to the NF. DF, divergent force field; NF, null field.

Finally, why does the CNS use the difficult and computationally costly strategy of controlling the full impedance geometry rather than the simple strategy of co-contracting all muscles? The particular shape of endpoint stiffness learned in the DF, large in the direction of instability and small in the direction of movement, suggests a trade-off between stability and high muscle activation. Increasing impedance enhances the robustness to external perturbations, but co-contraction of muscles increases metabolic cost¹². Moreover, motor output variability increases with muscle activation^{25,26}. Therefore, reducing the impedance will decrease both the metabolic cost and movement error. For these reasons, it is desirable to optimize impedance. The control problem faced by the CNS is to optimize the magnitude, shape and orientation of impedance to achieve stability while minimizing metabolic cost. □

Methods

Nine subjects performed the experiment (24–34 years of age; two females and seven males). The learning analysis used only data from the five naive subjects. The institutional ethics committee approved the experiments and the subjects gave informed consent before participation. The robot used is a parallel-link direct-drive air and magnet floating manipulandum (PFM), which moves in the horizontal plane. Details of its design and operation have been described previously^{19,21}. The PFM was set up by T. Yoshioka and H. Gomi. Subjects sat in a chair with a harness to constrain the trunk so that the elbow and shoulder joints could only move in the horizontal plane (two degrees of freedom). The forearm and wrist were held in a thermoplastic splint rigidly attached to the manipulandum. Therefore, subjects could not vary the kinematics of the arm to change its impedance. Subjects began by performing 100 movements in the NF, followed by 100–300 learning trials in the DF. Subjects then made 100 more movements in the DF, 20 of which had the force field removed unexpectedly (called after-effects). All movements were recorded during these sessions, including those not reaching the target. The start position, target (diameter 2.5 cm) and instantaneous hand position were displayed on a horizontal screen slightly above the arm. The prescribed movement time of 600 (±100) ms was indicated by acoustic signals.

Adaptation to the DF was quantified by calculating the error relative to a straight line joining the start position and target centre. The hand-path error E represents the area between the actual movement path and the straight line, and was calculated from the start time 0 (75 ms before crossing a hand-velocity threshold of 0.05 m s⁻¹), to the termination time T (when curvature exceeded 0.07 m⁻¹):

$$E = \int_0^T |x(t)| |y'(t)| dt \quad (2)$$

To test whether learning occurred, we compared the hand-path error in the first five learning trials and the last five learning trials using an analysis of variance (ANOVA) with random factor subjects. An ANOVA was also used to compare the last five trials in the NF and the last five trials of the learning phase. The presence of after-effects was tested by performing an ANOVA comparing the hand-path error of the last 20 movements in the NF to 20 after-effects trials.

Before each stiffness measurement session, the subjects were retrained in the force field to ensure that they had remembered the field (NF, 40 trials; DF, 80 trials). Eighty movements were then recorded in the force field, of which 40, selected randomly, incorporated displacements. These displacements, at the midpoint of each movement, were randomly applied in one of eight directions (0°, 45°, 90°, 135°, 180°, 225°, 270° or 315°). The endpoint stiffness was estimated from the resulting mean change in hand force and position over a 60-ms interval during the perturbation, starting from 120 ms after onset of the perturbation¹⁹.

Linear regression was carried out to determine whether the change in stiffness for movements in the DF relative to those in the NF could be explained by the corresponding change in force. By definition, the difference between torque produced in response to a perturbation $[dq_s \quad dq_e]^T$ in the DF and NF is given by:

$$\begin{bmatrix} \delta\tau_s \\ \delta\tau_e \end{bmatrix} = \Delta\mathbf{K}_q \begin{bmatrix} dq_s \\ dq_e \end{bmatrix} \quad (3)$$

where $\Delta\mathbf{K}_q$ is the difference between DF and NF joint stiffness, and $\delta\tau_s$ and $\delta\tau_e$ correspond to the difference between DF and NF shoulder and elbow torque owing to elastic resistance λ , respectively.

It has been shown that the joint stiffness is linearly dependent on the joint torque in posture¹⁵. Making the same assumption for movement, we obtain the linear relation:

$$\Delta\mathbf{K}_q = \begin{bmatrix} \alpha_1\Delta\tau_s & \alpha_3\Delta\tau_e \\ \alpha_3\Delta\tau_e & \alpha_4\Delta\tau_e \end{bmatrix} \quad (4)$$

where $\alpha_1 \dots \alpha_4$ are constants, and $\Delta\tau_s$ and $\Delta\tau_e$ are the difference between DF and NF shoulder and elbow driving torque measured by the force sensor λ . The difference in endpoint force can be obtained from equation (3) using the jacobian. Restricting the analysis to the x direction, since endpoint stiffness in the DF was principally modified in that direction, we then have:

$$\Delta F_x = [\Delta\tau_s dq_s \quad \Delta\tau_e dq_e \quad \Delta\tau_e dq_s] \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} \quad (5)$$

where γ_1 , γ_2 and γ_3 are constants that incorporate the α_i and the coefficients of the jacobian. R^2 values were computed for each subject using a linear system of 40 equations (with γ_1 , γ_2 and γ_3 as unknowns) representing the 40 trials where stiffness was measured. An R^2 value close to one would indicate that endpoint stiffness in the x direction was completely determined by the joint torques required to perform the task. An R^2 value close to zero would indicate that in every trial endpoint stiffness was controlled independently of the required joint torques. A more detailed derivation of these equations is available as Supplementary Information.

Received 8 June; accepted 27 September 2001.

- Shadmehr, R. & Mussa-Ivaldi, F. A. Adaptive representative of dynamics during learning of a motor task. *J. Neurosci.* **14**, 3208–3224 (1994).
- Lackner, J. R. & Dizio, P. Rapid adaptation to Coriolis force perturbations of arm trajectory. *J. Neurophysiol.* **72**, 299–313 (1994).
- Shadmehr, R. & Holcomb, H. H. Neural correlates of motor memory consolidation. *Science* **277**, 821–825 (1997).
- Krakauer, J. W., Ghilardi, M. F. & Ghez, C. Independent learning of internal models for kinematic and dynamic control of reaching. *Nature Neurosci.* **2**, 1026–1031 (1999).
- Kawato, M. Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* **9**, 718–727 (1999).
- Thoroughman, K. A. & Shadmehr, R. Learning of action through adaptive combination of motor primitives. *Nature* **407**, 742–747 (2000).
- Wolpert, D. M., Miall, R. C. & Kawato, M. Internal models in the cerebellum. *Trends Cogn. Sci.* **2**, 338–347 (1998).
- Colgate, J. E. & Hogan, N. Robust control of dynamically interacting systems. *Int. J. Control.* **48**, 65–88 (1988).
- McIntyre, J., Mussa-Ivaldi, F. A. & Bizzi, E. The control of stable postures in the multi-joint arm. *Exp. Brain Res.* **110**, 248–264 (1996).
- Rack, P. M. H. In *Handbook of Physiology* Sect. 1. *The Nervous System* (ed. Brooks, V. B.) 229–256 (American Physiological Society, Bethesda, 1981).
- Milner, T. E. & Cloutier, C. Compensation for mechanically unstable loading in voluntary wrist movement. *Exp. Brain Res.* **94**, 522–532 (1993).
- Hogan, N. The mechanics of multi-joint posture and movement control. *Biol. Cybern.* **52**, 315–331 (1985).
- Wong, J. & Hogan, N. Stability properties of human reaching movements. *Exp. Brain Res.* **107**, 125–136 (1995).
- Hunter, I. W. & Kearney, R. E. Dynamics of human ankle stiffness: variation with mean ankle torque. *J. Biomech.* **15**, 747–752 (1982).
- Gomi, H. & Osu, R. Task-dependent viscoelasticity of human multi-joint arm and its spatial characteristics for interaction with environments. *J. Neurosci.* **18**, 8965–8978 (1998).
- Mussa-Ivaldi, F. A., Hogan, N. & Bizzi, E. Neural, mechanical, and geometric factors subserving arm posture in humans. *J. Neurosci.* **5**, 2732–2743 (1985).
- Akazawa, K., Milner, T. E. & Stein, R. B. Modulation of reflex EMG and stiffness in response to stretch of human finger muscle. *J. Neurophysiol.* **49**, 16–27 (1983).
- De Serres, S. J. & Milner, T. E. Wrist muscle activation patterns and stiffness associated with stable and unstable mechanical loads. *Exp. Brain Res.* **86**, 451–458 (1991).
- Burdet, E. et al. A method for measuring endpoint stiffness during multi-joint arm movements. *J. Biomech.* **33**, 1705–1709 (2000).
- Lacquaniti, F., Carrozzo, M. & Borghese, N. A. Time-varying mechanical behavior of multi-jointed arm in man. *J. Neurophysiol.* **69**, 1443–1464 (1993).
- Gomi, H. & Kawato, M. Human arm stiffness and equilibrium-point trajectory during multi-joint movement. *Biol. Cybern.* **76**, 163–171 (1997).
- Milner, T. E., Cloutier, C., Leger, A. B. & Franklin, D. W. Inability to activate muscles maximally during co-contraction and the effect on joint stiffness. *Exp. Brain Res.* **107**, 293–305 (1995).
- Burdet, E., Osu, R., Franklin, D. W., Milner, T. E. & Kawato, M. in *Proc. 1999 ASME Annul. Symp. on Haptic Interfaces and Virtual Environments for Teleoperator Systems (DSC7C-7)* 421–428 (ASME Press, New York, 1999).
- Gomi, H. & Kawato, M. Task-dependent stiffness of human multi-joint arm during point-to-point movement. Technical Report ISRL-95-4 (NTT Basic Research Laboratories, Atsugi, 1995).
- Schmidt, R. A., Zelaznik, H., Hawkins, B., Frank, J. S. & Quinn, J. T. Jr Motor-output variability: a theory for the accuracy of rapid motor acts. *Psychol. Rev.* **47**, 415–451 (1979).
- Slikin, A. B. & Newell, K. M. Noise, information transmission, and force variability. *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 837–851 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

The experiments were performed at ATR. We thank A. Illiesch and T. Flash for discussions. We thank A. Smith, S. Scott, D. Wolpert, Z. Ghahramani, and S. Schaaf for giving valuable comments on an earlier version of the manuscript. This research was supported by the Special Coordination Fund for promoting Science and Technology of the Science and Technology agency of the Japanese Government; the Swiss National Science Foundation; the Natural Sciences and Engineering Research Council of Canada; and the Human Frontier Science Program.

Correspondence and requests for materials should be addressed to M.K. (e-mail: kawato@atr.co.jp). Further information about this work can be found at <http://www.bioeng.nus.edu.sg/research.htm>.

Genome sequence and gene compaction of the eukaryote parasite *Naucleoconium*

Michaël D. Katinka*, Simone Duprat*, Emmanuel Cornillot†, Guy Méténier‡, Fabienne Thomar†, Gérard Prensier†, Valérie Barbe*, Éric Peyretailade†, Philippe Brottier*, Patrick Wincker*, Frédéric Delbac†, Hicham El Alaoui†, Pierre Peyret†, William Saurin*, Manolo Gouy‡, Jean Weissenbach* & Christian P. Vivarès†

* Genoscope, UMR CNRS 8030, CP 5706, 91057 Evry cedex, France

† Parasitologie Moléculaire et Cellulaire, Laboratoire de Biologie des Protistes, UMR CNRS 6023, Université Blaise Pascal, 63177 Aubière cedex, France

‡ Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Université Lyon I, 69622 Villeurbanne cedex, France

Microsporidia are obligate intracellular parasites infesting many animal groups¹. Lacking mitochondria and peroxysomes, these unicellular eukaryotes were first considered a deeply branching protist lineage² that diverged before the endosymbiotic event that led to mitochondria. The discovery of a gene for a mitochondrial-type chaperone^{3–5} combined with molecular phylogenetic data^{6–9} later implied that microsporidia are atypical fungi that lost mitochondria during evolution. Here we report the DNA sequences of the 11 chromosomes of the ~2.9-megabase (Mb) genome of *Encephalitozoon cuniculi* (1,997 potential protein-coding genes). Genome compaction is reflected by reduced intergenic spacers and by the shortness of most putative proteins relative to their eukaryote orthologues. The strong host dependence is illustrated by the lack of genes for some biosynthetic pathways and for the tricarboxylic acid cycle. Phylogenetic analysis lends substantial credit to the fungal affiliation of microsporidia. Because the *E. cuniculi* genome contains genes related to some mitochondrial functions (for example, Fe–S cluster assembly), we hypothesize that microsporidia have retained a mitochondrion-derived organelle.

Encephalitozoon cuniculi infects various mammals, including humans, and can cause digestive and nervous clinical syndromes in HIV-infected or cyclosporine-treated people¹. Its reproduction proceeds as a sequence of two major stages: merogony, involving the multiplication of large, wall-lacking cells (meronts); and sporogony, leading to small, thick-walled spores. The sporadic invasive apparatus is characterized by a long polar tube that can be quickly extruded then used for transferring the sporoplasm into the target cell. Consisting of 11 linear chromosomes ranging from 217 to 315 kb, the *E. cuniculi* genome is remarkably reduced (~2.9 Mb)¹⁰. The nucleotide sequence of the smallest chromosome has been recently reported¹¹. The full sequencing of this minimal genome among eukaryotes was expected to provide insight into the metabolism and general biology of microsporidia and to help in the understanding of the evolutionary history of amitochondriate eukaryotes currently considered 'curious fungi'⁹ (see Table 1).

The chromosome sequences were determined through a plasmid library (~3 kb inserts) and a miniBAC library (~20–25 kb inserts) totalling 15 genome equivalents (~46 Mb). All chromosomes possess a 'unique sequences' core region flanked by two 28-kb divergently oriented regions, each including one ribosomal DNA unit^{11,12}. The subtelomeric repeats upstream from rDNA are mostly degenerated minisatellites, whereas downstream repeats consist essentially of non-polymorphic microsatellite arrays. The chromosome cores lack simple sequence repeats, minisatellite arrays and known transposable elements. No imprint of retrogenes or pseudoretrogenes of either polymerase (pol) II or pol III type was found. General features of genome organization are indicated in Table 1.

The gradual increase in G + C content at the core centre described in chromosome I (chrI)¹¹ exists in all the chromosomes (maximum 51.0%). The 1,997 protein-coding DNA sequences (CDSs) represent about 90% of the chromosome cores, as a result of generally short intergenic regions (see Supplementary Information). Gene density is slightly lower than that observed in the nucleomorph genome of the cryptomonad *Guillardia theta*¹³. Only about 44% of the CDSs are assigned to functional categories and about 6% to conserved hypothetical proteins (Fig. 1). In contrast to the nucleomorph genome¹³, no overlapping of CDSs with predicted functions was revealed. Structural or functional clusters are rare and never composed of more than two CDSs (for example, histones H3 and H4 on chrIX). Genome compaction can also be related to gene shortening, as indicated by the length distribution of all potential proteins (Fig. 2a). The mean and median lengths of all potential *E. cuniculi* proteins are only 359 and 281 amino acids, respectively. We compared the lengths of 350 proteins with *Saccharomyces cerevisiae* homologues (Fig. 2b). More than 85% of these proteins are shorter than in yeast, with a mean relative size difference of 14.6%. From the analysis of the protein size distributions derived from sequenced genomes, it has been suggested that the lengthening of proteins in eukaryotes (non-parasitic species) allows for more complex regulation networks¹⁴. Thus, protein shortening in *E. cuniculi* may reflect reduced protein–protein interactions as a result of various gene losses linked to the intracellular parasitic nature (Fig. 2a, b).

Perfect segmental duplications of 0.5–10-kb coding and non-coding sequences represent about 3.7% of the core region in average (from ~1% in chrII, III, IV and VI to ~7% in chrIX; see Supplementary Information). A segment carrying four enzyme-coding genes is perfectly duplicated in the extreme part of the chrI core¹¹ but partially duplicated near the end of chrVIII (truncated serine hydroxymethyltransferase gene). The genome core sequences exhibit a very low base polymorphism, restricted to a few positions in eight chromosomes. A duplicated gene homologous to CTP synthases (chrXI) revealed a rare polymorphic tract of 189 base pairs (bp) (116,847–117,036). Hybridization experiments indicate that each polymorphic sequence is present in about 50% of the DNA molecules (data not shown), suggesting that they are alleles and supporting the diploidy hypothesis^{10,12}.

A large proportion of CDSs is assigned to the conservation and transmission of genetic information as well as to protein modification and intracellular transport processes (Fig. 1), but with a significant degree of simplification, mainly related to the lack of DNA-containing organelles. For example, subunits for DNA pol I (δ), pol II (ε) and pol III (α but not β) are present while there is no candidate for the mitochondrial DNA polymerase γ. Apart from the telomerase catalytic subunit, no RNA-dependent reverse transcriptase was found. This and the lack of any retrotransposition elements may explain the absence of pseudoretrogenes. The potential transcription machinery includes RNA pol I, II and III, more than 70 messenger RNA transcription-associated proteins and a virtually

Table 1 General features of the *E. cuniculi* genome

Total sequenced length	2,507,519 bp
G + C content	
Protein-coding regions	47.6%
Intergenic regions	45.0%
Telomeric and subtelomeric regions	52.9%
No. of protein-coding sequences	1,997
Mean intergenic distance	129 bp
Gene density of chromosome cores	1 CDS per 1,025 bp
No. and sizes of spliceosomal introns	13 (23–52 bp)
No. of 16S–23S rRNA genes	22*
No. of 5S rRNA genes	3 (on chrV, VII, IX)
No. of tRNA genes	44†
No. and sizes of tRNA introns	2 (16, 42 bp)

* Two per chromosome.

† On all chromosomes.

complete set of genes for splicing, 5' and 3' processing. An (A + T)-rich consensus transcription initiation sequence is revealed in the 120-bp region upstream from the initiation codon of numerous genes, suggesting short 5' leaders. Seventy different eukaryotic-type ribosomal proteins (40 from the large subunit, 30 from the small subunit) are predicted, which is slightly lower than in the amitochondriate protist *Giardia lamblia* (74)¹⁵. Compared with the cytoplasmic ribosome of *S. cerevisiae*, the missing proteins are LP1, L14, L29, L38, L40, L41, S27 and S27A. Small putative spliceosomal introns with usual GT-AG boundaries were detected in 11 ribosomal protein genes (S8, S17, S24, S26, L5, L19, L27A, L37, L37A and L39). They start either in the initiator ATG or the next codon, as often observed in yeast¹⁶ or nucleomorph¹³ genomes. Two more internal introns create frame shifts within a CDP-diacylglycerol serine phosphatidyltransferase gene (chrXI). Two of the 44 transfer RNA genes (tDNA^{Ile} and tDNA^{Tyr}) also harbour a small intron.

Reduced metabolic capacities and low diversity of transporters can be inferred from the genome sequence, as illustrated in Fig. 3. The repertoire for the biosynthesis of amino acids is restricted to asparagine synthetase and serine hydroxymethyltransferase genes. Genes for *de novo* biosynthesis of purine and pyrimidine nucleotides are absent but several nucleotide interconversions are predicted. Genes encoding a fatty acid synthase complex are lacking, which supports the uptake of host-derived fatty acids¹⁷. The *E. cuniculi* spore membrane contains cholesterol. This sterol might also be of host origin, as no gene for the conversion of farnesyl-PP into cholesterol was detected. In contrast, *E. cuniculi* seems to be capable of synthesizing usual membrane phospholipids. Genes for principal enzymes for the synthesis and degradation of trehalose confirm that this disaccharide could be the major sugar reserve in microsporidia¹⁸, as in other fungi. A complete glycolytic glucose-to-pyruvate pathway is predicted. In contrast, genes required for the tricarboxylic acid cycle, fatty acid β -oxidation, respiratory electron-transport chain and the F_0F_1 -ATPase complex are absent. Thus, ATP production in microsporidia would be possible by substrate-level phosphorylation only. As proliferating microsporidia recruit host mitochondria near their plasma membrane, it has been proposed that these parasites could use host-derived ATP¹⁸. This is reinforced by the finding of four genes encoding ADP/ATP carrier proteins that are homologous to ADP/ATP translocases from chloroplasts and obligate intracellular bacteria (*Rickettsia* and *Chlamydia*) capable of importing host ATP¹⁹. The fate of pyruvate remains difficult to predict because of the lack of genes for lactate and ethanol fermentation as well as for glycolysis reversal. A potential cytosolic glycerol-3P dehydrogenase (GPDH) might serve to reoxidize the NADH produced during glycolysis. Surprisingly, two CDSs have significant similarity to the subunits of

the E1 component of the mitochondrial pyruvate dehydrogenase complex. Pyruvate decarboxylation could be inferred, but, in the absence of evidence for E2 and E3 components, a subsequent production of acetyl coenzyme A cannot be concluded.

Microsporidia have a presumably simplified Golgi apparatus in which a *cis-trans* polarity is not cytologically distinguishable but that is central in sporogony-specific secretion processes¹. The spore wall protein SWP1 (ref. 20) is encoded by a unique gene on chrX whereas two genes for the polar tube proteins PTP1 and PTP2 are arranged in tandem on chrVI²¹. The set of chaperones for protein folding includes a complete oligomeric TCP-1 complex, four members of the HSP70 system but no chaperonin CPN60. The mitochondrial-type HSP70 has been previously characterized in three different microsporidian species³⁻⁵. Initial steps of N-glycosylation using UDP-N-acetylglucosamine (UDP-GNAc) and GDP-mannose (GDP-Man) may occur, but further trimming by mannosidases associated with endoplasmic reticulum or Golgi apparatus and formation of a complex N-linked oligosaccharide are not supported. The major sugar used for O-linked glycosylations would be mannose (two mannosyltransferases of the fungal PMT family). The lack of genes for the two enzymes involved in phosphorylation of mannose residues argues for the absence of sorting of lysosomes. Membrane fusion and recognition of some target membrane processes are sustained by a restricted set of potential proteins for Golgi and post-Golgi trafficking. The constitutive secretion pathway leading to the plasma membrane may involve some characteristic Rab proteins. Several potential partners for *trans*-Golgi and endosome transport include β 1-adaptin, Vps1-like dynamin and vacuolar protein sorting-associated proteins, confirming that the Golgi-like apparatus is functionally polarized. Endocytosis of certain macromolecules was previously shown in *Spraguea lophii* sporoplasms¹⁸, when maintained *in vitro* in a cell culture medium. Several genes are also suggestive of an endocytosis pathway in *E. cuniculi*. This might drive the internalization of macromolecular ligands representing sources of fatty acids, cholesterol or iron (see Fig. 4).

The evolutionary origin of microsporidia has been much debated

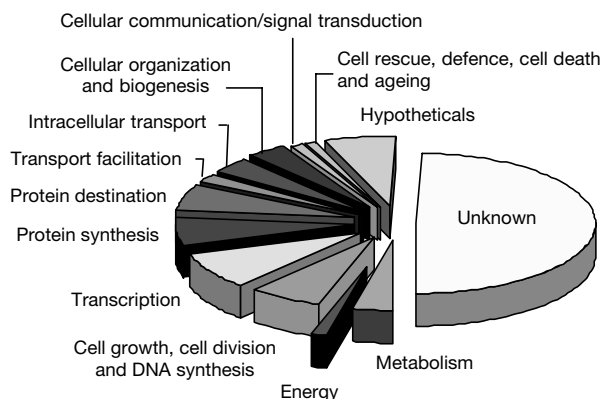


Figure 1 Distribution of predicted *E. cuniculi* proteins among functional groups. The corresponding gene list is given in the Supplementary Information.

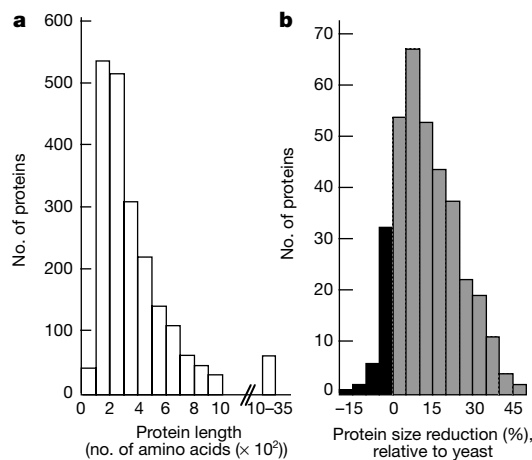


Figure 2 Sizes of *E. cuniculi* (*Ec*) proteins and comparison with *S. cerevisiae* (*Sc*) homologues. **a**, Distribution of the lengths of all the potential *E. cuniculi* protein chains. Only six have more than 2,000 amino acids (maximum 3,456 amino acids). **b**, Degrees of reduction in length of *E. cuniculi* proteins ($n = 350$) relative to those of *S. cerevisiae*, expressed as a percentage: $100(\text{Sc protein length} - \text{Ec protein length})/(\text{Sc protein length})$. The positive classes representative of shorter *E. cuniculi* proteins are in grey. The dynein heavy chain, the largest protein chain with a clearly predicted function (3,151 amino acids), is $\sim 23\%$ shorter than the yeast homologue (4,092 amino acids). Mean values associated with major functional categories are 7.3% for 'protein synthesis', 11.0% for 'protein destination', 11.4% for 'metabolism/energy', 14.4% for 'intracellular transport', 15.3% for 'transcription' and 20.1% for 'cell growth, cell division and DNA synthesis'.

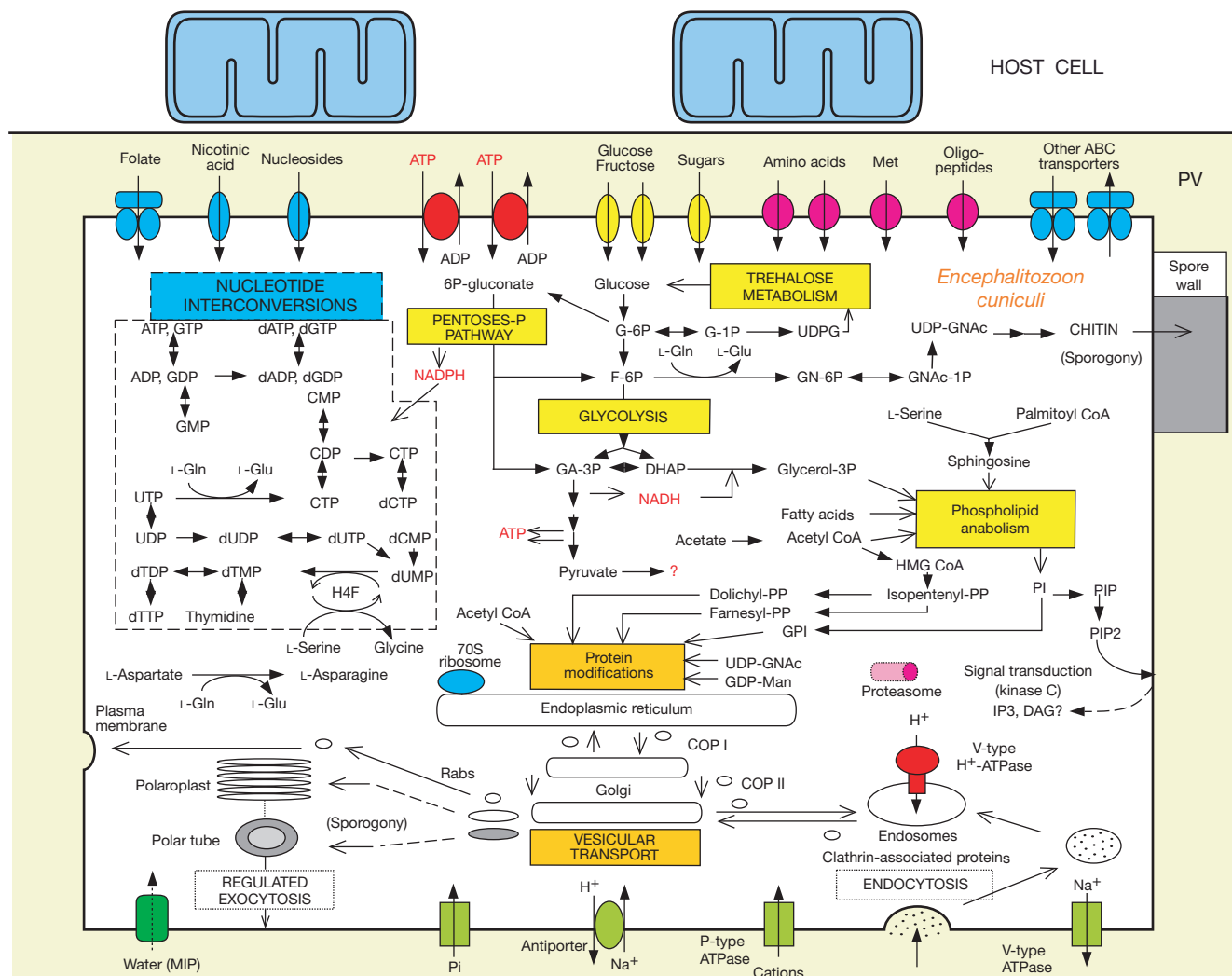
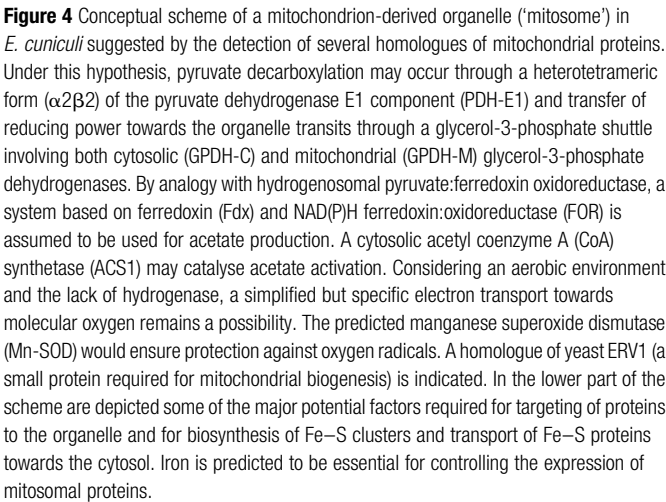


Figure 3 An overview of metabolism and transport in *E. cuniculi*, as deduced from genome sequence analysis. Pathways for nucleotide biosynthesis, energy production and chitin biosynthesis are indicated. Endocytosis and vesicular transport involving a *cis-trans* polarized Golgi apparatus are also illustrated. Potential transporters associated

with the plasma membrane are shown with indications on substrate specificity. Question marks correspond to major uncertainties about the fate of pyruvate and the production of second messengers for signal transduction. The parasite is represented within a parasitophorous vacuole (PV) of the host cytoplasm.

but strong evidence supporting a fungal origin of these organisms has recently accumulated⁶⁻⁹. The present genome sequence extends this evidence: phylogenetic analyses of putative genes for seryl-tRNA synthetase, transcription initiation factor IIB, subunit A of vacuolar ATPase, and a GTP-binding protein place microsporidia as a sister group of fungi with bootstrap supports ranging from 70% to 92% (see Supplementary Information; a systematic phylogenetic analysis of the genome will be presented elsewhere). Genes of putative mitochondrial evolutionary origin in the *E. cuniculi* genome were systematically sought by comparison with the 423 recently surveyed yeast mitochondrial proteins encoded by the nuclear and the mitochondrial genomes²². Twenty-two genes with significant similarity to the yeast genes were identified and phylogenetic analysis showed that six of them are closely related to homologues from α -proteobacteria, the bacterial group from which mitochondria are believed to derive²³. The yeast homologues of these six proteins are ATM1 (ABC transporter), ISU1/ISU2 (NIFU-like protein), NFS1 (unique homologue of bacterial ISC-S and NIF-S), SSQ1 (heat-shock protein of relative molecular mass 70,000), YAH1 (ferredoxin) and PDB1 (β -subunit of pyruvate dehydrogenase component E1). The first five proteins are typically involved in the Fe-S cluster assembly machinery, an essential function of mitochondria²⁴.

The presence of characteristic domains and key amino acids suggests that the potential mitochondrial-type proteins are functional. Moreover, PSORT analysis predicts amino-terminal presequences for the targeting of five of these proteins (see Supplementary Information). A common feature is an arginine residue at -2 relative to the cleavage site, similar to presequences of mitochondrial and hydrogenosomal proteins²⁵. The amitochondriate protozoan *Entamoeba histolytica* has recently been shown to contain a residual mitochondrion-derived organelle²⁶⁻²⁸ that some authors have called mitosome²⁸. Considering the set of potential *E. cuniculi* proteins usually associated with mitochondria, we propose that a cryptic organelle is also present in microsporidia (Fig. 4). This mitosome would be significantly different from hydrogenosomes found in several anaerobic unicellular eukaryotes (type II anaerobes)²⁹. Hydrogen production through pyruvate catabolism seems unlikely in microsporidia (because of a lack of a hydrogenase gene). The development of microsporidia in various aerobic host cells is suggestive of a rather high O₂ tolerance and therefore of an efficient protection against oxidative stress. No catalase gene is identified, in agreement with the lack of peroxisomes. Thus, in addition to glutathione and thioredoxin-based systems, *E. cuniculi* might use its unique manganese superoxide dismutase as an antioxidant.



This first report of the genome sequence of a eukaryotic parasite should stimulate proteomic approaches to identify gene products of interest for diagnosis and therapy of microsporidiosis, as well as to test the mitosome hypothesis. In addition, we expect the *E. cuniculi* genome to provide a useful reference for comparative genomics of microbial eukaryotes, particularly to identify the relative importance of shared genes among various evolutionarily distant intracellular parasites, including major human-infecting parasites such as *Plasmodium* and *Leishmania*. □

The reference mouse isolate of *Encephalitozoon cuniculi* (GB-M1), cloning and library construction, nucleotide sequencing, sequence validation and sequence analysis have been described in detail previously¹¹ (see Supplementary Information). Further information on recombinant DNA preparation, sequencing and sequence analysis can be found elsewhere³⁰. Annotation was manually performed with the help of AceDB and Artemis graphic interfaces. Genes were characterized by Glimmer prediction of coding DNA sequences, combined with BLAST all-homology results (BLAST X, BLAST N against ‘nr’, BLAST P against SwissProt, PSI-BLAST). Transfer RNA genes were detected with the tRNA Scan program. Spliceosomal-type introns were manually detected. Phylogenetic analyses were done as in Keeling *et al.*⁸. Gamma-corrected ML distances with eight rate categories and invariant sites were computed with TREE-PUZZLE version 5.0 and trees were derived with BioNJ. Bootstrapping was on 500 replicates with alpha parameters and

1. Wittner, M. & Weiss, L. M. *The Microsporidia and Microsporidiosis* (American Society of Microbiology, Washington DC, 1999).
2. Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vossbrinck, B. A. & Woese, C. R. Ribosomal RNA sequence suggests the microsporidia are extremely ancient eukaryotes. *Nature* **326**, 411–414 (1987).
3. Germot, A., Philippe, H. & Le Guyader, H. Evidence for loss of mitochondria in microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Mol. Biochem. Parasitol.* **87**, 159–168 (1997).
4. Hirt, R. P., Healy, B., Vossbrinck, C. R., Canning, E. U. & Embley, T. M. A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. *Curr. Biol.* **7**, 995–998 (1997).
5. Peyretailade, E. *et al.* Microsporidia, amitochondrial protists, possess a 70-kDa heat-shock protein gene of mitochondrial evolutionary origin. *Mol. Biol. Evol.* **15**, 683–689 (1998).
6. Hirt, R. P. *et al.* Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl Acad. Sci. USA* **96**, 580–585 (1999).
7. Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**, 972–977 (2000).
8. Keeling, P. J., Luker, M. A. & Palmer, J. D. Evidence from β -tubulin phylogeny that microsporidia evolved from within the Fungi. *Mol. Biol. Evol.* **17**, 23–31 (2000).
9. Van de Peer, Y., Ben Ali, A. & Meyer, A. Microsporidia: accumulating evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. *Gene* **246**, 1–8 (2000).
10. Biderre, C., Pagès, M., Méténier, G., Canning, E. U. & Vivarès, C. P. Evidence for the smallest nuclear genome (2.9 Mb) in the microsporidian *Encephalitozoon cuniculi*. *Mol. Biochem. Parasitol.* **74**, 229–231 (1995).
11. Peyret, P. *et al.* Sequence and analysis of chromosome I of the amitochondriate intracellular parasite *Encephalitozoon cuniculi* (Microspora). *Genome Res.* **11**, 198–207 (2001).
12. Brügère, J. F., Cornillot, E., Méténier, G., Bensimon, A. & Vivarès, C. P. *Encephalitozoon cuniculi* (Microspora) genome: physical map and evidence for telomere-associated rDNA units on all chromosomes. *Nucleic Acids Res.* **28**, 2026–2033 (2000).
13. Douglas, S. *et al.* The highly reduced genome of an enslaved algal nucleus. *Nature* **410**, 1091–1096 (2001).
14. Zhang, I. Protein-length distributions for the three domains of life. *Trends Genet.* **16**, 107–109 (2000).
15. Shirakura, T. *et al.* Characterization of the ribosomal proteins of the amitochondriate protist *Giardia lamblia*. *Mol. Biochem. Parasitol.* **112**, 153–156 (2001).
16. Spingola, M., Grate, L., Haussler, D. & Ares, M. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* **5**, 221–234 (1999).
17. El Alaoui, H., Bata, J., Bauchart, D., Doré, J.-C. & Vivarès, C. P. Lipids of three microsporidian species and multivariate analysis of the host–parasite relationship. *J. Parasitol.* **87**, 554–559 (2001).
18. Weidner, E., Findley, A. M., Dolgikh, V. & Skolova, J. In *The Microsporidia and Microsporidiosis* (eds Wittner, M. & Weiss, L. M.) 172–195 (American Society of Microbiology, Washington DC, 1999).
19. Wolf, Y. I., Aravind, L. & Koonin, E. V. Rickettsiae and Chlamydiae: evidence of horizontal transfer and exchange. *Trends Genet.* **15**, 173–175 (1999).
20. Böhne, W., Ferguson, D. J., Kohler, K. & Gross, U. Developmental expression of a tandemly repeated, glycine- and serine-rich spore wall protein in the microsporidian pathogen *Encephalitozoon cuniculi*. *Infect. Immun.* **68**, 2268–2275 (2000).
21. Delbac, F., Peuvrel, L., Méténier, G., Peyretailade, E. & Vivarès, C. P. Microsporidian invasion apparatus: identification of a novel polar tube protein and evidence for clustering of ptp1 and ptp2 genes in three *Encephalitozoon* species. *Infect. Immun.* **69**, 1016–1024 (2001).
22. Karlberg, O., Canback, B., Kurland, C. G. & Andersson, S. G. The dual origin of the yeast mitochondrial proteome. *Yeast* **17**, 170–187 (2000).
23. Andersson, S. G. *et al.* The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140 (1998).
24. Lill, R. & Kispal, G. Maturation of cellular Fe–S proteins: an essential function of mitochondria. *Trends Biochem. Sci.* **25**, 352–356 (2000).
25. Dyall, S. D. & Johnson, P. J. Origins of hydrogenosomes and mitochondria: evolution and organelle biogenesis. *Curr. Opin. Microbiol.* **3**, 404–411 (2000).
26. Rodriguez, M. A. *et al.* The pyruvate ferredoxin oxidoreductase is located in the plasma membrane and in a cytoplasmic structure in *Entamoeba*. *Microb. Pathog.* **25**, 1–10 (1998).
27. Mai, Z. *et al.* Hsp60 is targeted to a cryptic mitochondrion-derived organelle ('crypton') in the microaerophilic protozoan parasite *Entamoeba histolytica*. *Mol. Cell. Biol.* **19**, 2198–2205 (1999).
28. Tovar, J., Fischer, A. & Clark, C. G. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol. Microbiol.* **32**, 1013–1021 (1999).
29. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
30. Artgenave, F. *et al.* Genomic exploration of the hemiascomycetous yeasts: 2. Data generation and processing. *FEBS Lett.* **487**, 13–16 (2000).

We are grateful to everyone who provided computer resources and system management: D. Ferney and P. Bernard (LMA, UMR CNRS 6620), P. L. Reichstadt and B. Michel (LPC, IN2P3, UMR CNRS 6533). We thank B. Chebance and R. Guerry for technical assistance. L. Pereira, B. Duperrier, R. Météry and E. Delorme are acknowledged for their contribution to the development of the annotation environment in the Université Blaise Pascal. We also express gratitude to C. White for critical reading of the manuscript. F.D., H.E.A and E.P were supported by a grant from the Ministère de la Recherche.

Correspondence and requests for materials should be addressed to C.P.V. (e-mail: christian.vivares@lbp.univ-bpclermont.fr).

Innate antimicrobial peptide protects the skin from invasive bacterial infection

Victor Nizet*, Takaaki Ohtake†‡, Xavier Lauth†‡, Janet Trowbridge†‡, Jennifer Rudisill†‡, Robert A. Dorschner†‡, Vasumati Pestonjamas†‡, Joseph Piraino§, Kenneth Huttner§ & Richard L. Gallo*†‡

* Department of Pediatrics; and † Division of Dermatology, University of California, San Diego, California 92161, USA

‡ Veterans Affairs San Diego Healthcare System, San Diego, California 92161, USA

§ Division of Neonatology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

In mammals, several gene families encode peptides with antibacterial activity, such as the β -defensins and cathelicidins^{1–3}. These peptides are expressed on epithelial surfaces and in neutrophils, and have been proposed to provide a first line of defence against infection by acting as ‘natural antibiotics’^{4,5}. The protective effect of antimicrobial peptides is brought into question by observations that several of these peptides are easily inactivated^{6–8} and have diverse cellular effects that are distinct from antimicrobial activity demonstrated *in vitro*^{9–13}. To investigate the function of a specific antimicrobial peptide in a mouse model of cutaneous infection, we applied a combined mammalian and bacterial genetic approach to the cathelicidin antimicrobial gene family¹⁴. The mature human (LL-37)¹⁵ and mouse (CRAMP)¹⁶ peptides are encoded by similar genes (*CAMP* and *Cnlp*, respectively), and have similar α -helical structures, spectra of antimicrobial activity and tissue distribution. Here we show that cathelicidins are an important native component of innate host defence in mice and provide protection against necrotic skin infection caused by Group A *Streptococcus* (GAS).

To assess directly the function of cathelicidins *in vivo*, we generated mice that are null for *Cnlp* by targeted recombination. A targeting vector was constructed in which exons 3 and 4, encoding the entire mature domain of CRAMP, were replaced with PGK-neo flanked 5' by a genomic Xba/R1 fragment and 3' by a 5.5-kilobase (kb) R1 fragment of *Cnlp* (Fig. 1a). This construct was introduced into 129/SVJ embryonic stem cells by electroporation and subjected to G418 selection. Recombinant clones were screened by polymerase chain reaction (PCR) using a forward primer 5' to the deletion construct and reverse primer within the neomycin resistance gene. Embryonic stem cell clones were injected into C57BL/6 blastocysts and transferred to foster mothers. Chimaeric offspring were crossed with C57BL/6 females, whose heterozygous progeny were identified by PCR (Fig. 1b), and these were immediately backcrossed into 129/SVJ. Northern blot analysis confirmed the absence of CRAMP in *Cnlp*-null bone marrow (Fig. 1c). *Cnlp*-null mice had normal fetal development, were fertile, survived into adulthood, and demonstrated no obvious phenotype when housed under aseptic barrier-controlled conditions.

The cathelicidins CRAMP and LL-37 (refs 15, 16) are greatly increased in the skin after wounding, owing to their release from neutrophil granules and increased synthesis by keratinocytes¹⁷. We chose to use GAS for our mouse infection model because the injury accompanying this pathogen leads to a large increase in the local accumulation of CRAMP¹⁷. Moreover, GAS are highly sensitive to cathelicidin antimicrobial action, even under culture conditions that inactivate peptide killing of other bacteria frequently described as sensitive (for example, *Escherichia coli*)¹⁷. Subcutaneous injection of GAS in mice induces a necrotic lesion that is histopathologically similar to that seen with invasive human infections. After identical

injections of cathelicidin-sensitive GAS in wild-type, heterozygous and homozygous-null mice, CRAMP-deficient mice were observed to develop much larger areas of infection (Fig. 2a, b). Lesion areas increased more rapidly, reached larger maximal size, and persisted longer in CRAMP-deficient mice than in normal littermates while heterozygotes tended to have lesions of intermediate size (Fig. 2c). Cultures of equal amounts of tissue from lesions biopsied at day 7 after injection demonstrated persistent infection with β -haemolytic GAS in CRAMP-deficient mice but not in normal mice (Fig. 2d). No difference in GAS lesion size was seen when wild-type parental strains C57BL/6 and 129/SVJ were compared.

A complementary approach to demonstrating the importance of cathelicidins in host defence is to examine the effects *in vivo* of altering bacterial sensitivity to CRAMP. If the antimicrobial action of cathelicidin is essential to control a GAS skin infection, as suggested by the experiments with *Cnlp*-null mice, then CRAMP-resistant GAS should be more pathogenic than CRAMP-sensitive GAS in normal mice. A transposon mutant library was generated from wild-type GAS strain NZ131 by random integration of Tn917 into the bacterial chromosome. In contrast to the parent strain, bacterial growth was observed in a pooled library of transposon mutants exposed to increasing concentrations of the cathelicidin antimicrobial peptide. Southern blot analysis of several isolated colonies demonstrated clonality. The sequence flanking the single chromosomal integration of Tn917 in the cathelicidin-resistant mutant (NZ131-CR) was identified and compared to the recently completed GAS genome database¹⁸. The Tn917 insertion mapped to an open reading frame (GenBank AAK34584) encoding a predicted product of relative molecular mass 28,200 (M_r 28.2K) with the signature helix-turn-helix motif of the GntR family of bacterial transcription regulation proteins¹⁹ (Fig. 3a). To demonstrate that Tn917 disruption at this locus was reproducibly associated with an inducible cathelicidin-resistance phenotype, targeted plasmid integrational mutagenesis of the *gntR*-related open reading frame was

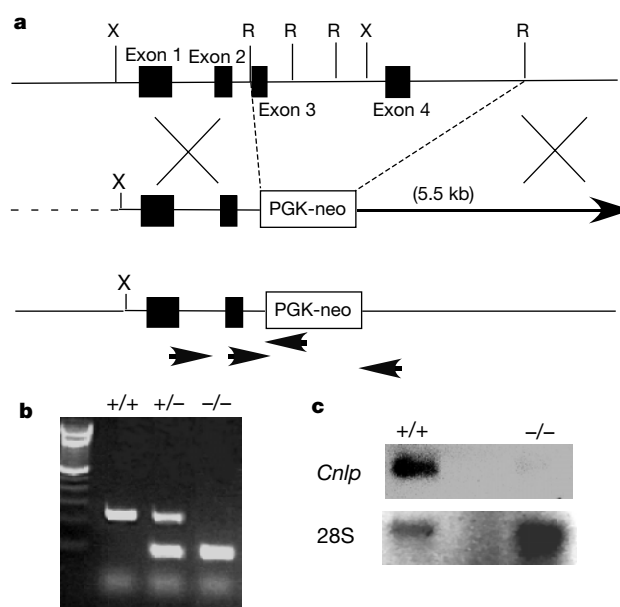


Figure 1 Disruption of the *Cnlp* gene encoding CRAMP in mice. **a**, Gene targeting strategy: structure and partial restriction map of *Cnlp*. Top line, filled boxes represent exons 1–4. X, *Xho*I; R, *Eco*RI. Second line, Targeting construct replaces exon 3–4 R fragment with PGK-neo and retains 5' X–R fragment and 5.5-kb 3' R fragment. Dotted lines, gene fragment replaced by PGK-neo; crosses, flanking regions in construct to target homologous recombination; bottom line, inactivated allele; arrowheads are PCR primers. **b**, PCR genotype results of tail DNA amplified with primers shown in **a**. **c**, Northern blot results of total RNA extracted from bone marrow. *Cnlp* probe is directed to exon 4 alone. 28S RNA is shown as a loading control.

performed in the parent GAS NZ131 (Fig. 3a). The putative cathelicidin-resistance regulatory gene was designated *crgR* (for cathelicidin resistance gene regulator) and the targeted plasmid integrational mutant was designated NZ131:*crgR*.KO. NZ131:*crgR*.KO exhibited inducible CRAMP-resistance to identical minimal inhibitory concentrations (MICs) when compared to the original Tn917 mutant (Fig. 3b). Comparison of the kinetics of bacterial killing upon exposure to CRAMP further demonstrated that mutant bacteria were resistant to the bactericidal action of the peptide (Fig. 3c). Both the transposon and targeted CRAMP-resistant mutants exhibited slight growth inhibition in enriched culture media compared to the parent GAS strain (data not shown), suggesting a metabolic cost to the bacteria associated with acquired antimicrobial peptide resistance and possible pleiotropic effects of this regulatory gene mutation. We note that mutations in transcriptional regulatory genes of *Salmonella*²⁰ and *Burkholderia pseudomallei*²¹ have also been associated with a change in susceptibility to cationic antimicrobial peptides.

CRAMP-resistant mutants of GAS were compared to wild-type GAS for their ability to produce necrotizing cutaneous infection in a normal mouse model. Mice infected with Tn917 mutant NZ131-CR or targeted mutant NZ131:*crgR*.KO had lesions of larger size and longer duration than those infected with the parent GAS strain (Fig. 3d–f). Biopsy and quantitative culture of the lesions from mice infected with the CRAMP-resistant mutants showed persistent infection at day 7 in contrast to the microbial clearing observed with the CRAMP-sensitive parent GAS. Thus, cathelicidin-resistant mutants demonstrated increased virulence *in vivo*. In effect, induction of cathelicidin resistance in the bacterial pathogen reproduced the phenotype of the CRAMP gene knockout in the mouse infection model.

At day 3, histological examination of GAS-induced skin lesions from CRAMP-deficient mice showed abundant neutrophilic infiltrates comparable to those seen in the wild-type mice (data not shown). No differences were seen in the circulating peripheral leukocyte morphology or count as determined by microscopic examination of blood smears (data not shown) or by fluores-

cence-activated cell sorting (FACS) analysis (Fig. 4a, b). Furthermore, leukocytes derived from CRAMP-deficient mice were functionally competent and similar to wild-type leukocytes in oxidative burst activity (Fig. 4c). These observations suggest that although CRAMP-deficient mice appear to have normal recruitment of an acute neutrophil inflammatory response, the absence of the antimicrobial peptide in the neutrophil granule and epidermal keratinocyte leads to defects in the control of GAS infection. The importance of CRAMP in phagocytic clearance of GAS was reflected in results of whole-blood killing assays (Fig. 4d). Fresh blood from wild-type mice reduced bacterial counts after incubation for one hour, whereas fresh blood from CRAMP-deficient mice allowed bacterial replication. Bacterial replication was also seen when CRAMP-resistant GAS was incubated in wild-type mouse blood. As expected, the CRAMP-resistant phenotype of GAS did not affect

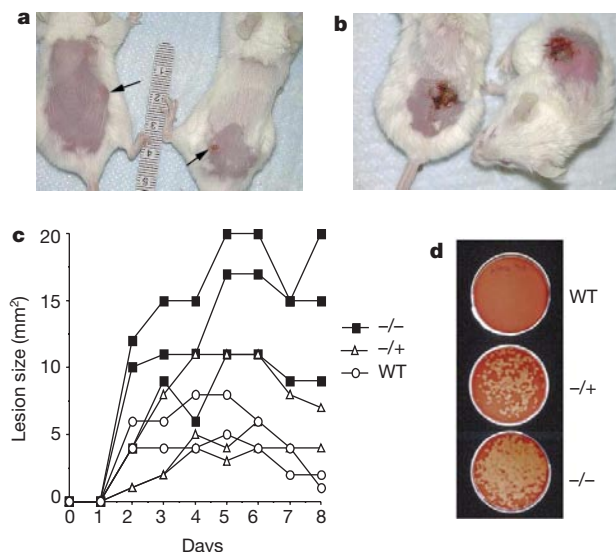


Figure 2 *Cnlp*-deletion renders mice susceptible to severe GAS infection. **a**, Wild-type (WT) and **b**, *Cnlp*-null mice following subcutaneous inoculation with GAS. Scale bar is in centimetres. **c**, Area of necrotic ulcer in individual WT (circle), CRAMP +/- (triangle), and CRAMP -/- (square) mice shown against days after infection. **d**, GAS bacteria cultured from tissue biopsies of WT, CRAMP +/- and CRAMP -/- mice. The experiment shown is representative of five studies in which similar differences were observed ($n = 12$, $P < 0.001$).

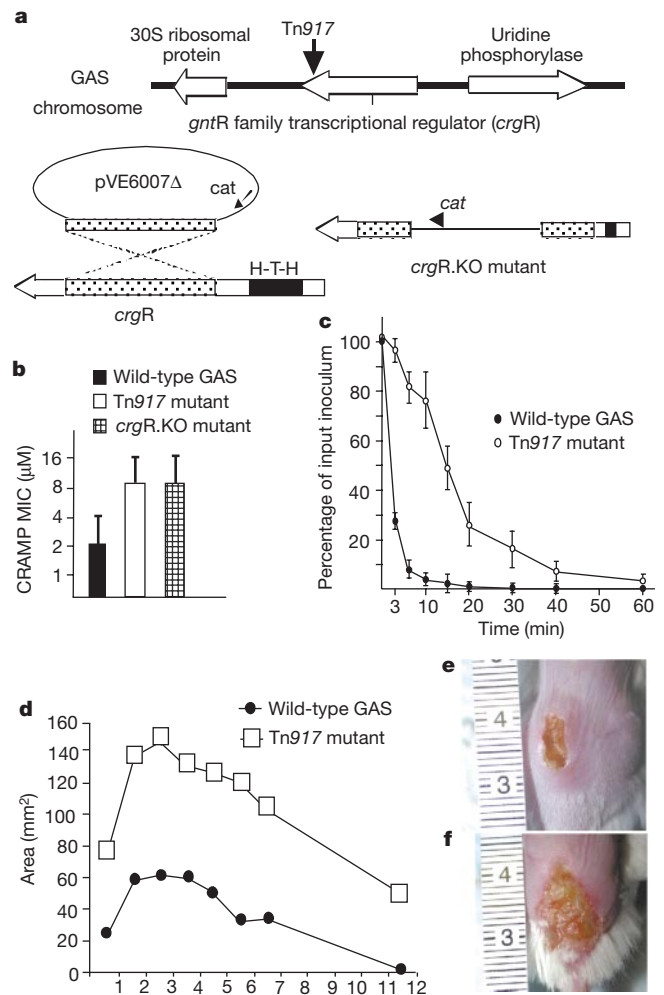


Figure 3 GAS mutants resistant to CRAMP produce more severe infections in normal mice. **a**, Chromosomal location of the Tn917 insertion in CRAMP-resistant GAS mutant NZ131:*crgR*.KO. H-T-H, helix-turn-helix; cat, chloramphenicol acetyltransferase. **b**, Minimum inhibitory concentrations of CRAMP against wild-type GAS parent strain NZ131 (black bar), Tn917 mutant NZ131-CR (white bar), and targeted plasmid integrational mutant NZ131:*crgR*.KO (hatched bar). **c**, Bacterial killing kinetics of 32 μM CRAMP versus wild-type GAS and Tn917 mutant NZ131-CR. **d**, Area of necrotic ulcer formation in wild-type mice infected with CRAMP-resistant mutant NZ131-CR (square) versus CRAMP-sensitive parent strain NZ131 (circle). Data are means of three mice and similar differences were observed in five experiments with NZ131-CR ($n = 10$, $P < 0.01$) and two experiments with NZ131:*crgR*.KO ($n = 8$, $P < 0.001$). Representative lesions in mice infected with **e**, CRAMP-sensitive parent strain NZ131 versus **f**, CRAMP-resistant mutant NZ131-CR. Scale is shown in centimetres.

the degree of bacterial replication in the blood of CRAMP-deficient mice. Similarly, CRAMP-deficient mice showed no difference in skin lesion size when injected with wild-type or CRAMP-resistant GAS (data not shown).

We have thus demonstrated *in vivo* that endogenous expression of a mammalian antimicrobial peptide provides defence against an invasive bacterial infection. Bacterial resistance to cathelicidins altered disease outcome in a similar fashion to elimination of host cathelicidin production. These paired findings suggest that the specific antimicrobial activity of the cathelicidin is itself necessary for bacterial clearance and innate skin immunity. Some earlier studies^{22–25} have also suggested *in vivo* protection by mammalian antimicrobial peptides. For example, inhibition of antimicrobial peptide activation leads to increased wound colonization by *Staphylococcus epidermidis* in pigs²², natural resistance to antimicrobial peptide action appears to contribute to *Salmonella* virulence in mice²³, downregulation of enteric cathelicidin and β -defensin-1 expression is correlated with *Shigella* infections in humans²⁴, and improved lung clearance of *Pseudomonas aeruginosa* is associated with overexpression of a human antimicrobial peptide gene in transgenic mice²⁵. These observations, combined with our own, suggest that the study of the effects of cathelicidin deficiency at other epithelial surfaces will further define the function of this family of antimicrobial peptides in disease. Other antimicrobial peptides (such as the defensins) that are co-expressed with cathelicidins act synergistically *in vitro*, and may have a similar role *in vivo*.

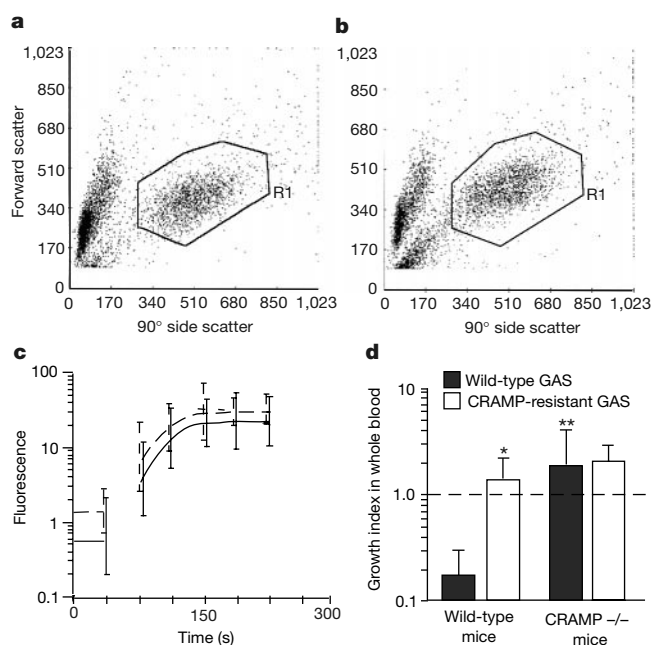


Figure 4 *Cnlp*-null mouse blood leukocytes are equivalent in relative number and oxidative burst capacity, but deficient in bacterial killing when compared to leukocytes derived from wild-type mice. Flow cytometry of leukocytes derived from **a**, *Cnlp*-null mice and **b**, wild-type mice. Granulocyte population is indicated (R1). Data are representative of the results obtained for seven individuals per genotype. The proportion of leukocytes, monocytes and lymphocytes were not significantly different by analysis with paired *t*-test. **c**, Oxidative burst capacity as measured by fluorescence before (0–40 s) and after (~75–250 s) the addition of a reagent designed to stimulate leukocyte phagocytic and oxidative activity as described in Methods. The mean log fluorescence per time is indicated for wild-type (dashed line) and *Cnlp*-null (solid line) mice with error bars indicating the range of >95% of fluorescent readings collected at each time point. Data are representative of three individual experiments. **d**, GAS killing activity of whole mouse blood. Data are expressed as growth index (c.f.u. bacteria after 1 h incubation in blood versus initial bacterial c.f.u.) and represent the mean of four mice tested in each group. Asterisk, $P < 0.02$; double asterisk, $P < 0.01$ versus normal mouse/wild-type GAS.

Furthermore, the influence of cathelicidins and defensins on cell-mediated immunity through leukocyte recruitment^{26,27} may also contribute to their defense properties. Finally, analysis of the bacterial genetic basis of cathelicidin sensitivity can shed light on the mechanism(s) of action of this class of antimicrobial peptides and the reasons why resistance has not emerged in the evolution of certain pathogenic bacterial species. We believe a combined genetic approach, studying the susceptibility factors of both host and pathogen, to be fruitful in understanding these important components of the innate immune system. □

Methods

Generation of *Cnlp*-deficient mice

Cnlp-null mice were generated by targeted disruption of *Cnlp* through homologous recombination. Embryonic stem cell culture and blastocyst injection were performed at Genome Systems (St Louis, Missouri), now Incyte Genomics (Palo Alto, California). Screening for positive recombinant embryonic stem cells was performed by PCR and confirmed by Southern blot. Selected embryonic stem cell clones were karyotyped and two normal clones were used for germline transmission. Male chimaeras were bred with C57BL/6 females and germline mice were identified using PCR analysis with the following *Cnlp* primers: (5'-CCAGGGACTTCCATCCAGTAGAC-3', 5'-TGTTTCTCAGATCCTTGGGAGC-3', 5'-AATTTTCTTGAACCGAAAGGGC-3') and neo reverse primer (5'-AGACTGCCTTGGGAAAAGCG-3'). Correct insertion was confirmed using a primer 5' to the Xba site at the start of the targeting construct (5'-GTCAGGTCACATAGGCAATGGG-3') and the reverse neo primer. The identity of PCR products was confirmed by direct sequencing. Heterozygote offspring from chimaeric matings were backcrossed into 129/SVJ for two generations.

Bacterial mutagenesis

GAS strain NZ131 is a well characterized T14/M49 patient isolate that produces the virulence factors streptolysin S and O, and pyrogenic exotoxin (SPE) B²⁸. NZ131 was randomly mutated with Tn917 using the temperature-sensitive delivery vector pTV1OK (ref. 28). Mutant clone NZ131-CR was identified by Southern analysis following serial passage of the pooled library in increasing concentrations of CRAMP (4–16 μ M). Single-primer PCR²⁹ with a Tn917-specific outward primer was used to identify the NZ131-CR mutation site, and a BLAST (<http://www.ncbi.nlm.nih.gov/BLAST>) homology search against the GAS genome (see <http://www.genome.ou.edu/strep.html>)¹⁸ revealed identity with the open reading frame we have designated *crgR*. For targeted mutagenesis of *crgR*, an intragenic fragment was amplified by PCR and cloned into the temperature-sensitive vector pVE6007Δ. The resultant 'knockout' plasmid was introduced into NZ131 by electroporation, and chloramphenicol (Cm)-resistant transformants identified at the permissive temperature for plasmid replication (30 °C). Single-crossover Campbell-type chromosomal insertion was selected by shifting to the non-permissive temperature (37 °C) while maintaining Cm selection; fidelity of the site-directed recombination event and disruption of the targeted open reading frame was confirmed by PCR²⁸.

Antimicrobial assays

For MIC determination, serial dilutions of the peptide were made in H₂O and 10 μ l of each concentration added to replicate wells of a 96-well flat-bottom tissue-culture plate. GAS were grown in Todd Hewitt Broth (THB, Difco) to early log phase (absorbance at 600 nm, $A_{600} = 0.2$). Bacteria were diluted in THB to $A_{600} = 0.001$ ($\sim 2 \times 10^5$ colony-forming units (c.f.u.) ml⁻¹). In triplicate, 90 μ l of this bacterial suspension was added to each well containing antimicrobial peptide. Growth was monitored at A_{600} through overnight incubation at 37 °C, and MIC was calculated as the lowest peptide concentration yielding no detectable growth. For antimicrobial killing kinetics, $\sim 2 \times 10^5$ c.f.u. of each strain were exposed to 32 μ M CRAMP and incubated at 37 °C. Dilutions were plated at time points from 3 to 60 min for determination of surviving c.f.u.

Mouse model of GAS infection

Invasiveness of GAS in mouse skin was measured by modification of a previously described GAS infection model³⁰. Procedures were approved by the Veterans Affairs (VA) San Diego Healthcare System subcommittee on animal studies. For *Cnlp*-null mice experiments, the backs of sex-matched adult littermates were shaved and hair removed by chemical depilation (Nee) then injected subcutaneously with 50 μ l of a mid-logarithmic growth phase ($A_{600} = 0.6$, $\sim 5 \times 10^7$ c.f.u.) of GAS NZ131 complexed to Cytodex beads as a carrier. For experiments with CRAMP-resistant GAS, 12-week-old Balb/c female mice were similarly prepared and injected with equal numbers of NZ131, NZ131-CR, or NZ131:*crgR*.KO. Lesion sizes were measured daily and statistical significance between groups evaluated by multiple regression analysis on VassarStats (<http://faculty.vassar.edu/lowry/VassarStats.html>).

Flow cytometry

Some 200–500 μ l of whole blood was collected into 12 ml cold phosphate-buffered saline (PBS) heparin (2 U ml⁻¹) (Sigma number H3393). Cells were centrifuged and erythrocytes lysed using Becton Dickinson FACS Lysis Buffer (number 349202). Unlysed cells were washed once with 1 ml DMEM 2% BSA (Sigma number A2153), then resuspended in

200–400 μ l 4% paraformaldehyde (PFA) (number S898-07 JT Baker). Approximately 2×10^5 cells were evaluated for forward and side scatter characteristics by flow cytometry under the direction of J. Nordberg at the VA Core Flow Facility to determine the relative percentage of lymphocytes, monocytes and granulocytes. For oxidative burst assay, blood was collected and lysed as above then resuspended at 4 °C in approximately 200 μ l endotoxin and pyrogen-free PBS without Ca^{2+} and Mg^{2+} (number 70013-032, Gibco BRL) containing 5 mM glucose (Sigma number G5146). Before the oxidative burst assay, 200 μ l of PBS at 37 °C containing 1.5 mM Mg^{2+} and 1.0 mM Ca^{2+} was added to the cell suspension. The Fc OxyBURST Green Assay Reagent (Molecular probes number F-2902) was used to evaluate cell uptake and oxidative burst activity. Background fluorescence was recorded, then oxidative burst reagent added to a concentration of 30 $\mu\text{g ml}^{-1}$ (5 μ l).

Whole blood killing assay

Fresh mouse blood was collected as above and 35 μ l added to 10 μ l of THB containing $(1-2) \times 10^5$ c.f.u. of freshly diluted log-phase GAS. The mixture was incubated for 1 h at 37 °C with gentle agitation, and dilutions of the mixture plated on agar plates for enumeration of c.f.u. The growth index was calculated as the ratio of bacterial c.f.u. recovered versus the initial bacterial inoculum. Student's *t*-test analysis was performed using the Microsoft Excel statistical package.

Received 28 June; accepted 1 October 2001.

1. Scott, M. G. & Hancock, R. E. Cationic antimicrobial peptides and their multifunctional role in the immune system. *Crit. Rev. Immunol.* **20**, 407–431 (2000).
2. Boman, H. G. Innate immunity and the normal microflora. *Immunol. Rev.* **173**, 5–16 (2000).
3. Ganz, T. & Lehrer, R. Antibiotic peptides from higher eukaryotes: biology and applications. *Mol. Med. Today* **5**, 292–297 (1999).
4. Ganz, T. *et al.* Defensins: natural peptide antibiotics of human neutrophils. *J. Clin. Invest.* **76**, 1427–1435 (1985).
5. Zasloff, M. Magainins, a class of antimicrobial peptides from *Xenopus* skin: Isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proc. Natl Acad. Sci. USA* **84**, 5449–5453 (1987).
6. Goldman, M. J. *et al.* Human β -defensin-1 is a salt-sensitive antibiotic in lung that is inactivated in cystic fibrosis. *Cell* **88**, 553–560 (1997).
7. Wang, Y., Agerberth, B., Lofgren, A., Almstedt, A. & Johansson, J. Apolipoprotein A-I binds and inhibits the human antibacterial/cytotoxic peptide LL-37. *J. Biol. Chem.* **273**, 33115–33118 (1998).
8. Travis, S. M. *et al.* Bactericidal activity of mammalian cathelicidin-derived peptides. *Infect. Immun.* **68**, 2748–2755 (2000).
9. Gallo, R. L. *et al.* Syndecans, cell surface heparan sulfate proteoglycans, are induced by a proline-rich antimicrobial peptide from wounds. *Proc. Natl Acad. Sci. USA* **91**, 11035–11039 (1994).
10. Shi, J., Ross, C., Leto, T. & Blecha, F. PR-39, a proline-rich antibacterial peptide that inhibits phagocyte NADPH oxidase activity by binding to Src homology 3 domains of p47^{phox}. *Proc. Natl Acad. Sci. USA* **93**, 6014–6018 (1996).
11. Lencer, W. *et al.* Induction of epithelial chloride secretion by channel-forming cryptidins 2 and 3. *Proc. Natl Acad. Sci. USA* **94**, 8585–8589 (1997).
12. Rizzo, A., Zanetti, M. & Gennaro, R. Cytotoxicity and apoptosis mediated by two peptides of innate immunity. *Cell. Immunol.* **189**, 107–115 (1998).
13. Chertov, O. *et al.* Identification of defensin-1, defensin-2, and CAP37/azurocidin as T-cell chemoattractant proteins released from interleukin-8-stimulated neutrophils. *J. Biol. Chem.* **271**, 2935–2940 (1996).
14. Zanetti, M., Gennaro, R., Scocchi, M. & Skerlavaj, B. Structure and biology of cathelicidins. *Adv. Exp. Med. Biol.* **479**, 203–218 (2000).
15. Gudmundsson, G. H. *et al.* The human gene *Fall39* and processing of the cathelin precursor to the antibacterial peptide LL-37 in granulocytes. *Eur. J. Biochem.* **238**, 325–332 (1996).
16. Gallo, R. L. *et al.* Identification of CRAMP, a cathelin-related antimicrobial peptide expressed in the embryonic and adult mouse. *J. Biol. Chem.* **272**, 13088–13093 (1997).
17. Dorschner, R. A. *et al.* Cutaneous injury induces the release of cathelicidin antimicrobial peptides active against group A *Streptococcus*. *J. Invest. Dermatol.* **117**, 91–97 (2001).
18. Ferretti, J. J. *et al.* Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc. Natl Acad. Sci. USA* **98**, 4658–4663 (2001).
19. Peekhaus, N. & Conway, T. Positive and negative transcriptional regulation of the *Escherichia coli* glucuronate regulon gene *gntT* by GntR and the cyclic AMP (cAMP)–cAMP receptor protein complex. *J. Bacteriol.* **180**, 1777–1785 (1998).
20. Miller, S. I., Pulkkinen, W. S., Selsted, M. M. & Mekalanos, J. J. Characterization of defensin resistance phenotypes associated with mutations in the *phoP* virulence regulon of *Salmonella typhimurium*. *Infect. Immun.* **58**, 3706–3710 (1990).
21. Burtnick, M. N. & Woods, D. E. Isolation of polymyxin B-susceptible mutants of *Burkholderia pseudomallei* and molecular characterization of genetic loci involved in polymyxin B resistance. *Antimicrob. Agents Chemother.* **43**, 2648–2656 (1999).
22. Cole, A. M. *et al.* Inhibition of neutrophil elastase prevents cathelicidin activation and impairs clearance of bacteria from wounds. *Blood* **97**, 297–304 (2001).
23. Groisman, E. A., Parra-Lopez, C., Salcedo, M., Lippes, C. J. & Heffron, F. Resistance to host antimicrobial peptides is necessary for *Salmonella* virulence. *Proc. Natl Acad. Sci. USA* **89**, 11939–11943 (1992).
24. Islam, D. *et al.* Downregulation of bactericidal peptides in enteric infections: A novel immune escape mechanism with bacterial DNA as a potential regulator. *Nature Med.* **7**, 180–185 (2001).
25. Bals, R., Weiner, D., Meegalla, R. & Wilson, J. Transfer of a cathelicidin peptide antibiotic gene restores bacterial killing in a cystic fibrosis xenograft model. *J. Clin. Invest.* **103**, 1113–1117 (1999).
26. De, Y. *et al.* LL-37, the neutrophil granule- and epithelial cell-derived cathelicidin, utilizes formyl peptide receptor-like 1 (FPR1) as a receptor to chemoattract human peripheral blood neutrophils, monocytes, and T cells. *J. Exp. Med.* **192**, 1069–1074 (2000).
27. Yang, D. *et al.* β -defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. *Science* **286**, 525–528 (1999).
28. Nizet, V. *et al.* Genetic locus for streptolysin S production by group A streptococcus. *Infect. Immun.* **68**, 4245–4254 (2000).

29. Karlyshev, A. V., Pallen, M. J. & Wren, B. W. Single-primer PCR procedure for rapid identification of transposon insertion sites. *Biotechniques* **28**, 1078–1082 (2000).
30. Betschel, S., Borgia, S., Barg, N., Low, D. & De Azavedo, J. Reduced virulence of group A streptococcal Tn916 mutants that do not produce streptolysin S. *Infect. Immun.* **66**, 1671–1679 (1998).

Acknowledgements

This work was supported by grants from the NIH (R.L.G. and V.N.), a VA merit award (R.L.G.), and grants from the American Skin Association (R.L.G.) and the Rockefeller Brothers Foundation (V.N.).

Correspondence and requests for materials should be addressed to R.L.G. (e-mail: rgallo@vapop.ucsd.edu).

The E2F1–3 transcription factors are essential for cellular proliferation

Lizhao Wu*, Cynthia Timmers*, Baidehi Maiti*, Harold I. Saavedra*, Ling Sang*, Gabriel T. Chong*, Faison Nuckolls†, Paloma Giangrande†, Fred A. Wright*, Seth J. Field‡, Michael E. Greenberg‡, Stuart Orkin§, Joseph R. Nevins†, Michael L. Robinson|| & Gustavo Leone*

* Division of Human Cancer Genetics, Department of Molecular Virology, Immunology and Medical Genetics, and Department of Molecular Genetics; and || Division of Molecular and Human Genetics, Children's Research Institute, The Ohio State University, Columbus, Ohio 43210, USA

† Department of Genetics, Howard Hughes Medical Institute, Duke University Medical Center, Durham, North Carolina 27710, USA

‡ Department of Neuroscience, § Howard Hughes Medical Institute, Children's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

The retinoblastoma tumour suppressor (Rb) pathway is believed to have a critical role in the control of cellular proliferation by regulating E2F activities^{1,2}. E2F1, E2F2 and E2F3 belong to a subclass of E2F factors thought to act as transcriptional activators important for progression through the G1/S transition³. Here we show, by taking a conditional gene targeting approach, that the combined loss of these three E2F factors severely affects E2F target expression and completely abolishes the ability of mouse embryonic fibroblasts to enter S phase, progress through mitosis and proliferate. Loss of E2F function results in an elevation of p21^{Cip1} protein, leading to a decrease in cyclin-dependent kinase activity and Rb phosphorylation. These findings suggest a function for this subclass of E2F transcriptional activators in a positive feedback loop, through down-modulation of p21^{Cip1}, that leads to the inactivation of Rb-dependent repression and S phase entry. By targeting the entire subclass of E2F transcriptional activators we provide direct genetic evidence for their essential role in cell cycle progression, proliferation and development.

The delineation of a pathway controlling the progression of cells out of quiescence, through G1 and into S phase, has been established^{1,2}. Principal events in this pathway include the activation of cyclin-dependent kinases (CDKs), the coordinated phosphorylation of Rb and p130 by cyclin–CDK complexes, and the subsequent release and accumulation of E2F activities^{1,2}. Although E2F has an essential role in control of cell growth during *Drosophila* development^{4,5}, current knockout mouse models have failed to demonstrate a similar requirement for any E2F family member in mammals^{6–12}. One interpretation of these observations is that under normal circumstances, loss of a single E2F member can be functionally compensated by other related E2F activities.

Of the six known E2F family members, E2F1, E2F2 and E2F3 can specifically interact with Rb, and their expression is cell-cycle regulated^{13,14}. To test for functional redundancy among this subclass

of E2F family members, we generated and interbred *E2F1*, *E2F2* and *E2F3* mutant mice^{6,9} (Fig. 1), and found that although *E2F1*^{-/-}*E2F2*^{-/-} mice were viable and developed to adulthood, *E2F1*^{-/-}*E2F3*^{-/-} and *E2F2*^{-/-}*E2F3*^{-/-} animals died early during embryonic development (Supplementary Information Table 1), at or just before 9.5 embryonic days (E9.5), pointing to a central role for *E2F3* in mouse development.

To explore directly the potential role for this subclass of E2F transcription activators in cellular proliferation, we measured the ability of cells deficient for these E2F family members to proliferate. We introduced a conditional or floxed *E2F3* allele (*E2F3*^{flf}; Fig. 1) into *E2F* mutant backgrounds and obtained *E2F1*^{-/-}*E2F3*^{flf}, *E2F2*^{-/-}*E2F3*^{flf} and *E2F1*^{-/-}*E2F2*^{-/-}*E2F3*^{flf} embryos at the predicted frequencies (Supplementary Information Table 1 and data not shown), confirming that the floxed *E2F3* allele did not have any adverse effect on the development of animals. Infection of mouse embryonic fibroblasts (MEFs), which were derived from these embryos, with a retrovirus expressing the Cre recombinase resulted in the specific deletion of the floxed exon 3 (Fig. 1d) and complete loss of both E2F3a and E2F3b proteins¹⁵ (Fig. 2c, top panel). As shown in Fig. 2a, the differences in the growth rates between control- and Cre-treated wild-type cells were insignificant ($P = 0.455$), suggesting that ectopic Cre expression from our retroviral vectors does not affect cell growth to any appreciable extent. However, Cre-mediated ablation of *E2F3* from *E2F3*^{flf} MEFs resulted in a slight decrease in their rate of proliferation relative

to control retrovirus-infected cells ($P = 0.024$). Moreover, co-culture of control- and Cre-infected *E2F3*^{flf} MEFs eventually led to the outgrowth of non-deleted *E2F3*^{flf} cells (data not shown). These findings, together with the observation that *E2F3*^{-/-} MEFs proliferate slower than their wild-type counterparts⁸ (Supplementary Information Fig. 1), suggest a cell-autonomous role for *E2F3* in the control of cellular proliferation.

Whereas cells deficient for *E2F1* and *E2F2* proliferated robustly, Cre-mediated ablation of *E2F3* from *E2F1*^{-/-}*E2F3*^{flf} or *E2F2*^{-/-}*E2F3*^{flf} MEFs severely impaired proliferation (Fig. 2a; $P = 0.002$ for both). Notably, loss of the three E2F family members abolished any measurable proliferation ($P \ll 0.0001$). A statistical test designed to account for serial correlation (see Methods) confirmed that the cell growth defect is due to Cre-mediated inactivation of *E2F3*. We also determined the long-term growth potential of MEFs deficient for the various E2F family members. Consistent with data presented above, loss of *E2F3* reduced efficiency of colony formation by approximately 50%; additional loss of *E2F1* or *E2F2* accentuated this reduction and loss of all three E2F family members almost completely abolished any colony-forming ability (Fig. 2b). Most colonies arising from Cre-treated *E2F1*^{-/-}*E2F3*^{flf} or *E2F2*^{-/-}*E2F3*^{flf} cells were deleted for exon 3, as determined by polymerase chain reaction (PCR) genotyping of individual colonies (Fig. 2e, top panel). In contrast, the few colonies that did arise from Cre-treated *E2F1*^{-/-}*E2F2*^{-/-}*E2F3*^{flf} cells possessed at least one *E2F3* floxed allele (Fig. 2e, bottom panel; 79 colonies pooled from multiple experiments were analysed, $P \ll 0.0001$), supporting the idea that E2F activators are essential for cellular proliferation. We conclude that

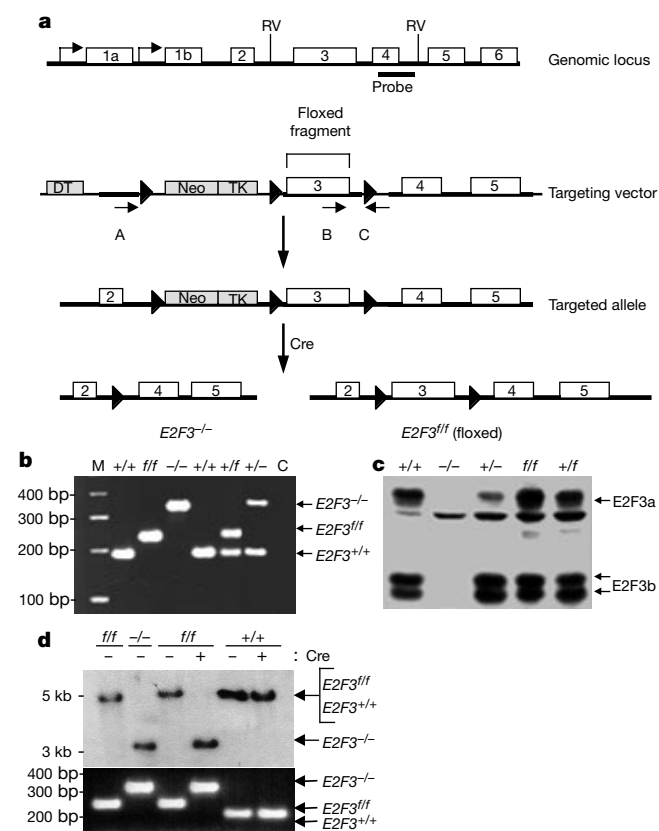


Figure 1 Targeted disruption of the *E2F3* locus. **a**, Two alternative splicing forms (*E2F3a* and *E2F3b*) with separate promoters (bent arrows) are encoded by the *E2F3* locus¹⁵. The solid bar represents a *SpeI*–*EcoRV* fragment used as a Southern probe. *loxP* sites are indicated as solid triangles; thin arrows represent PCR primers. DT, diphtheria toxin; TK, thymidine kinase; RV, *EcoRV*. **b**, PCR analysis of mice with the indicated genotypes. M, DNA marker; C, negative control. **c**, Western blot of MEFs with the indicated genotypes using an *E2F3*-specific antibody. **d**, Southern blot (top panel) and PCR (bottom panel) of control- or Cre-retrovirus-infected *E2F3*^{flf} MEFs. Wild-type (+/-) and *E2F3*^{-/-} (-/-) MEFs are indicated.

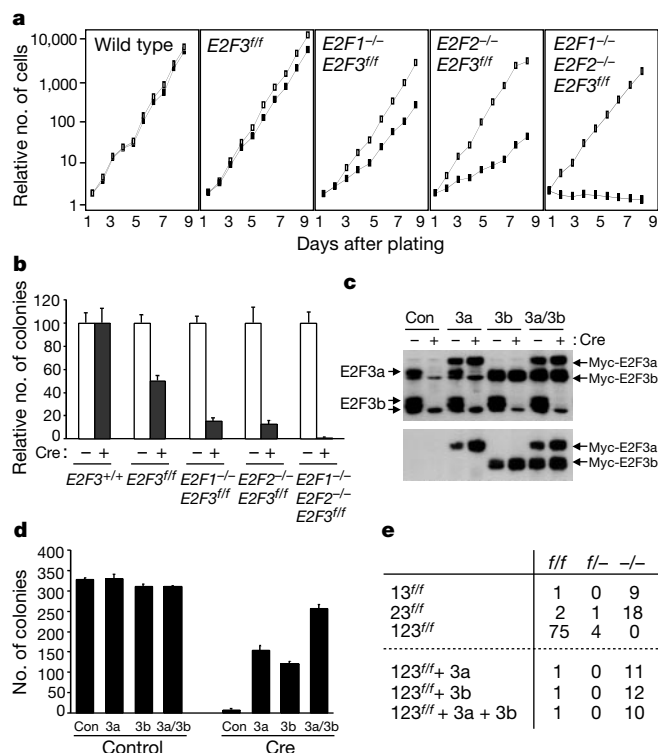


Figure 2 E2Fs are required for cellular proliferation. **a**, Growth curves of MEFs infected with a control (open boxes) or a Cre retrovirus (filled boxes). **b**, Colony formation assay of MEFs infected with a control (-) or a Cre retrovirus (+). **c**, Western blots using an antibody against *E2F3* (top panel) or Myc epitope (bottom panel). *E2F1*^{-/-}*E2F2*^{-/-}*E2F3*^{flf} MEFs were co-infected with the indicated retroviruses. -, pBpuro; +, pBpuro-Cre; Con, pBhygro and pBblea; 3a, pBhygro-myc-tagged *E2F3a*; 3b, pBblea-myc-tagged *E2F3b*. **d**, Colony formation assays of *E2F1*^{-/-}*E2F2*^{-/-}*E2F3*^{flf} MEFs that were co-infected with the indicated retroviruses. **e**, Number of colonies obtained for the indicated genotypes from experiments shown in **b** and **d**.

E2F1 and E2F2 cooperate with E2F3 to sustain normal proliferation and suggest that the severe growth defect resulting from their loss leads to the early embryonic lethality of $E2F1^{-/-}E2F3^{-/-}$ or $E2F2^{-/-}E2F3^{-/-}$ animals.

To confirm that loss of E2F3 was the primary factor responsible for the lack of growth observed on Cre treatment of $E2F1^{-/-}E2F2^{-/-}E2F3^{+/+}$ cells, we assessed the ability of Myc-tagged versions of $E2F3a$ and $E2F3b$ to rescue the growth defect of triple-knockout (TKO) cells. Western blot analysis demonstrated that the levels of ectopically expressed E2F3 proteins were similar to those of endogenous E2F3 normally found in wild-type or $E2F3$ -floxed cells (Fig. 2c). Reintroduction of E2F3a or E2F3b partially restored the colony-forming ability of TKO cells, and the combined expression of both proteins almost completely rescued the defect in colony formation (Fig. 2d). Furthermore, PCR genotyping confirmed that in cells where E2F3a and/or E2F3b was reintroduced, most of the emerging colonies were deleted for exon 3 (Fig. 2e), demonstrating that the rescue was due to the ectopic expression of E2F3 proteins and not to a failure in the Cre-mediated ablation of E2F3.

The proliferation assays described above suggest a particularly important function for E2F3 in growth control, but also show significant contributions made by E2F1 and E2F2 towards the control of cellular proliferation. These data, however, can not

distinguish between a model where individual E2F transcription factors might execute specific functions, such as the regulation of unique sets of gene targets, or a model where each E2F would contribute towards a pool of total E2F activity. To address this issue we sought to rescue the growth phenotype of TKO cells with haemagglutinin (HA)-tagged versions of each E2F activator. The results shown in Supplementary Information Fig. 2 indicate that E2F2 could only partially rescue the growth defect of TKO cells. Western blot assays indicated that E2F2 and E2F3 were expressed to equivalent levels, but E2F1 was consistently expressed at lower levels (Supplementary Information Fig. 2). The inherent ability of E2F1 to induce apoptosis probably led to the selection of cells with low expression of E2F1, making the comparison of their growth properties with those cells expressing exogenous E2F2 or E2F3 difficult to evaluate. Although some redundancy in their function is evident, our current data support a model where each E2F activator contributes unique functions towards the control of cell cycle and proliferation. The replacement of specific E2Fs by other family members using knock-in strategies *in vivo* might provide more definitive answers to this issue of functional specificity among E2F family members.

We next addressed whether the reduced growth rates of E2F mutant MEFs could be attributed to a specific defect in the G1/S transition. MEFs treated with either the Cre-expressing or control virus were synchronized by serum deprivation, then stimulated to proliferate by the addition of serum, and assayed for E2F-target gene activation and S-phase entry. Although peak levels of E2F target genes were not markedly affected by loss of E2F3 alone, their normal serum-induced accumulation was delayed by 2–3 h, a delay that was augmented further by the additional loss of either E2F1 or E2F2 (Supplementary Information Fig. 3 and data not shown). Cells deficient for all three E2F activators failed to induce many, but not all, E2F target genes in response to serum, including *Dhfr*, *Cdc6*, *Mcm3*, thymidine kinase (*Tk*) and DNA polymerase α (*DNA pol\alpha*) (Fig. 3). Moreover, the low levels of these transcripts in unsynchronized cell populations correlated with their inability to be induced by serum (Fig. 3a, both lanes labelled P). Notably, other putative E2F targets, including the *cyclin E* gene, were consistently elevated in quiescent TKO cells, and their levels remained relatively unaffected in unsynchronized cell populations (Fig. 3a, top panel). This finding is consistent with these targets being regulated mainly by Rb-mediated repression. Finally, reintroduction of E2F3 into TKO cells restored the normal activation of E2F target genes that occurs after mitogenic induction of quiescent cells (Fig. 3b). The levels of exogenous E2F3 protein introduced into these cells were nearly physiological, as indicated by western blot assays and its inability to activate E2F targets under conditions of low serum (Figs 2c and 3b, time 0). These findings support the idea that E2F activators are critical factors important for cell-cycle-dependent gene expression.

Cre-mediated loss of E2F3 from each of the E2F3-floxed cell lines impaired S phase entry, as measured by incorporation of 5-bromodeoxyuridine (BrdU) (Fig. 4a), to an extent that correlated with the severity of the proliferation and target activation defects described above. The defect of cell cycle entry of TKO cells is not due to their general inability to respond to growth signals, as the growth-factor-induced activation of the mitogen-activated protein kinase kinase and phosphatidylinositol-3-OH kinase immediate-early pathways remained unaffected by the loss of E2F family members (Supplementary Information Fig. 3c). These findings indicate that E2F members grouped in the activator subclass, namely E2F1, E2F2 and E2F3, function at the G1/S transition to regulate entry into S phase and thereby promote cellular proliferation.

To determine whether the loss of E2F function would lead to an accumulation of cells in a specific phase of the cell cycle, TKO cells were analysed for DNA content by flow cytometry. As shown in Fig. 4b, control virus-treated $E2F1^{-/-}E2F2^{-/-}E2F3^{+/+}$ cells, rendered

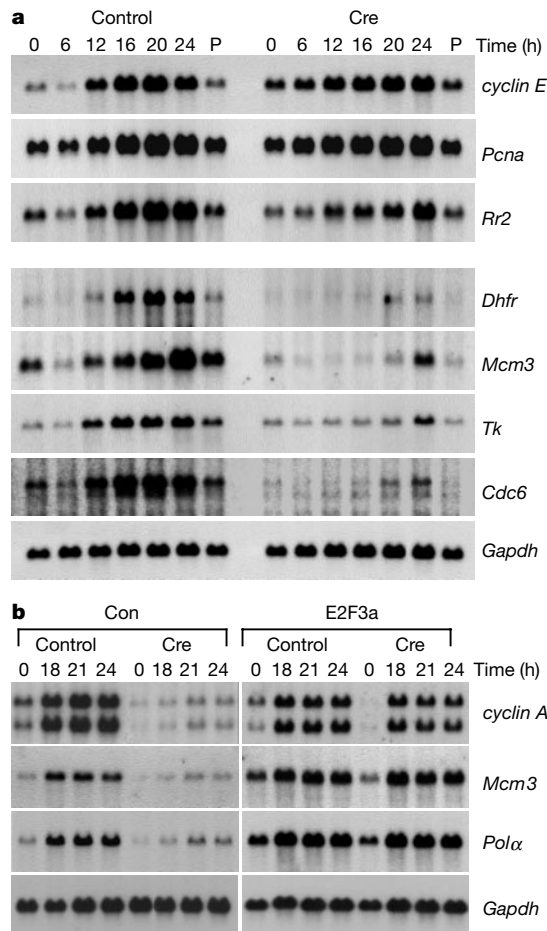


Figure 3 E2F activators regulate G1/S-specific target gene activation. **a**, Northern blot analysis of various E2F-responsive genes in $E2F1^{-/-}E2F2^{-/-}E2F3^{+/+}$ MEFs that were infected with a control or a Cre retrovirus. Serum-starved MEFs were induced to re-enter the cell cycle by addition of serum. RNA was extracted from serum-starved cells and serum-stimulated cells at different time points. P, proliferating cells. **b**, Northern blot analysis of various E2F-responsive genes in $E2F1^{-/-}E2F2^{-/-}E2F3^{+/+}$ MEFs that were treated in **a**. Con, pBHygro; E2F3a, pBHygro-myc-tagged E2F3a. Gapdh was used as a probe to verify equal loading.

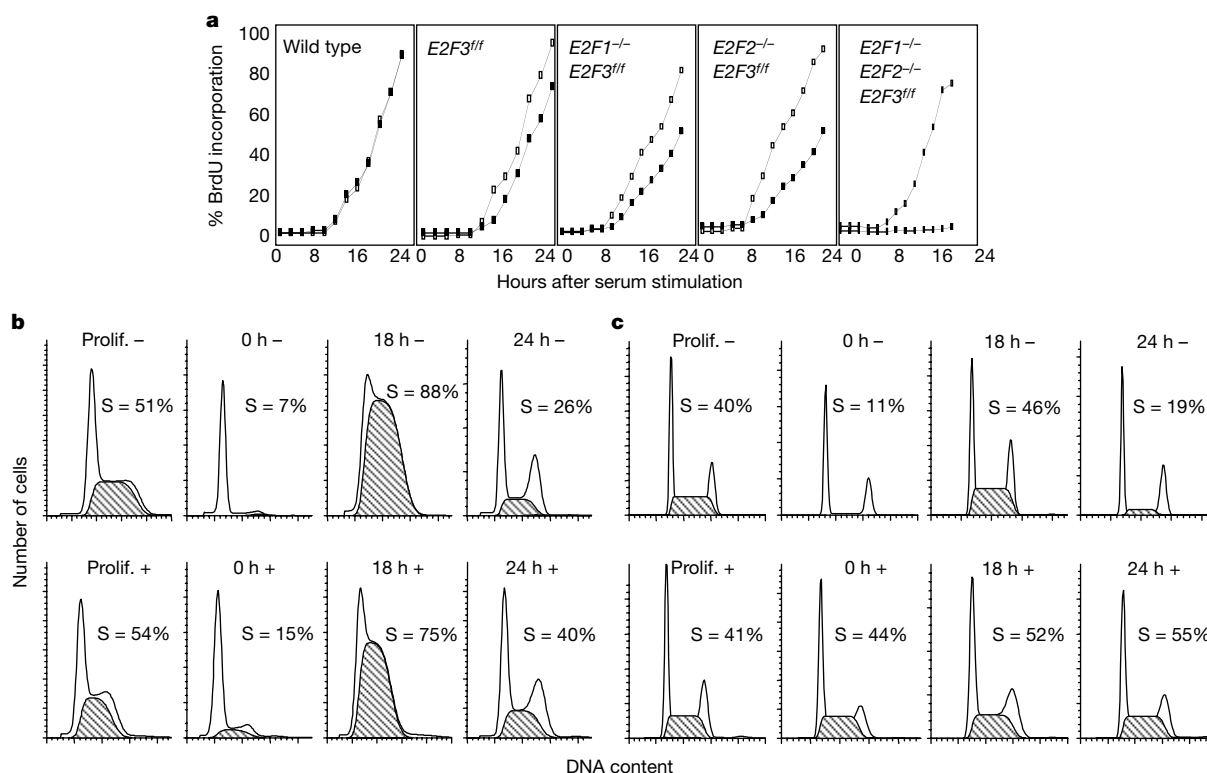


Figure 4 E2Fs are required for S phase entry and cell cycle progression. **a**, BrdU incorporation assays of MEFs that were infected with a control (open boxes) or a Cre retrovirus. **b, c**, Cell cycle analysis by FACS for wild-type MEFs (**c**) or *E2F1^{-/-}E2F2^{-/-}E2F3^{fl/fl}* MEFs (**c**) that were infected with a control (top panels (-)) or a Cre retrovirus (bottom

panels (+)). In **c**, a fraction of cells had become tetraploid by the time the experiment was performed, as indicated by the appearance of two distinct peaks in serum-starved samples (0 h). Hatched areas indicate the percentage of cells (S) containing an S-phase content of DNA.

quiescent by serum deprivation, accumulated a G1 content of DNA, and could be efficiently induced to enter the cell cycle after serum addition, as suggested by the timely accumulation of an S phase content of DNA. The similar profiles of control- and Cre-treated wild-type cells indicated that expression of Cre itself had little or no effect on the ability of cells to proliferate or progress through the cell cycle (Fig. 4b). Unsynchronized TKO cell populations had a similar profile of DNA content as control cells, except that most of these cells were also negative for BrdU incorporation (Fig. 4c, compare top and bottom samples; and data not shown). Moreover, TKO cells failed to accumulate a G1 content of DNA on serum deprivation and failed to respond to serum addition (Fig. 4c). Together, these findings show that ablation of all three E2F family members arrests cell growth irrespective of cell cycle position, indicating a role for these E2Fs throughout the cell cycle. Recent work also suggests a continuing role for E2F during S phase that seems to be important for the regulation of mitotic regulators and progression of cells through G2/M¹⁶. This view is consistent with recent genome-wide analyses of cell-cycle-regulated genes, where a large fraction of E2F-regulated gene targets were found to encode proteins known to be involved in the progression of cells into G2 and through mitosis^{17,18}.

To identify relevant downstream activities that might be responsible for mediating the growth arrest observed upon loss of E2F3, we assessed the status of various inhibitors of the cell cycle. Whereas p53 or p27^{Kip1} remained unchanged, p21^{Cip1} protein levels were markedly elevated in TKO cells (Fig. 5a). This increase in p21^{Cip1} protein could be accounted for by a concomitant increase in its levels of messenger RNA (Fig. 5b). Consistent with these findings, Cdk2-associated activities were markedly reduced and Rb was found to be predominantly in its hypophosphorylated state (Fig. 5a, c). Furthermore, cyclin-B-associated kinase activity was also significantly reduced in TKO cells, even though a large

population of the cells possessed either an S phase or G2 content of DNA. These results might, at least in part, explain the apparent growth arrest that occurs through various phases of the cell cycle, namely in G1, S and G2/M of TKO cells. The fact that overexpression of cyclin E/Cdk2 could drive TKO cells into S phase (Fig. 5d) suggests that the upregulation of p21^{Cip1} protein is probably a relevant consequence resulting from the ablation of E2F1, E2F2 and E2F3.

The status of Rb-E2F-associated DNA binding activities were also assessed in these cells. Consistent with previous studies, electrophoretic mobility shift assays (EMSA) of lysates from control- or Cre-treated *E2F1^{-/-}E2F2^{-/-}E2F3^{fl/fl}* cells synchronized by serum deprivation, revealed the characteristic Rb-related p130-E2F4/5 protein complexes¹⁴. These G0-specific complexes are thought to recruit histone deacetylase and remodelling activities, and are presumed to mediate transcriptional repression of E2F target genes¹⁹⁻²⁵. In contrast to control-treated MEFs, serum stimulation of TKO cells did not lead to the dissociation and disappearance of p130-E2F4/5 complexes nor to the appearance of the S-phase-specific p107-E2F4 complex (Fig. 5e). These findings are consistent with the observed increase in p21^{Cip1} protein and decrease in Cdk activity and Rb phosphorylation (Fig. 5a, c). We propose that E2F activators mediate cell cycle progression by two mechanisms: first, they can directly activate cell-cycle-dependent gene expression, and second, they can activate a positive feedback loop, through inhibition of p21^{Cip1} protein accumulation, that would promote Rb phosphorylation and complete derepression of Rb-E2F target genes. The decrease in E2F target gene expression observed in TKO cells (Fig. 3a) could be viewed to result from both the continued Rb-mediated transcriptional repression and the absence of E2F-mediated transcription activation. Although our results establish the critical nature of E2F1, E2F2 and E2F3 in the control

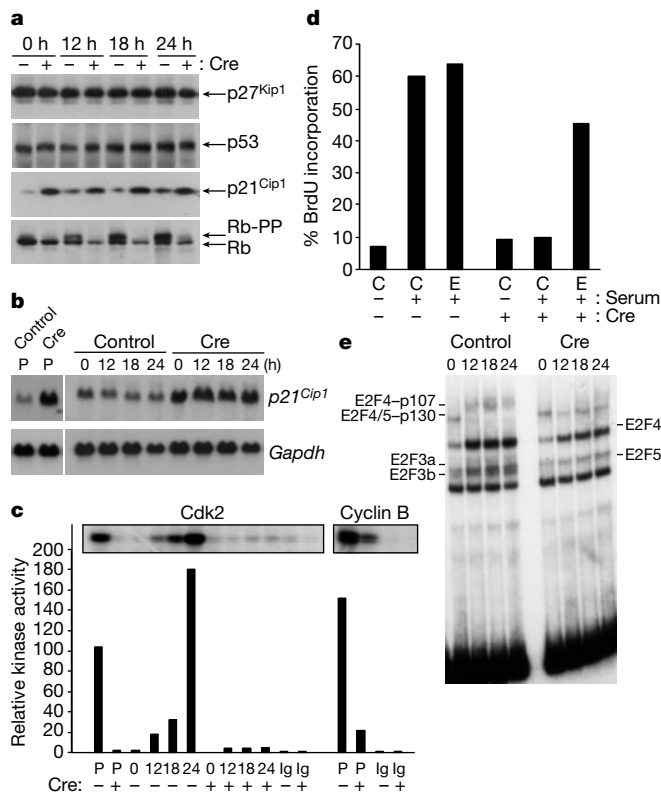


Figure 5 Loss of E2Fs leads to an elevation of p21^{Cip1}. **a**, Western blots of control- or Cre-retrovirus-infected *E2F1^{-/-}E2F2^{-/-}E2F3^{+/+}* MEFs using antibodies against the indicated cell cycle regulators. Rb-PP, phosphorylated form of Rb. **b**, p21 Northern blot of *E2F1^{-/-}E2F2^{-/-}E2F3^{+/+}* MEFs that were treated as in **a**. P, proliferating conditions. **c**, Kinase assays of *E2F1^{-/-}E2F2^{-/-}E2F3^{+/+}* MEFs treated as in **a**. Ig, rabbit immunoglobulin-γ as controls. The graph represents the relative kinase activity. **d**, BrdU incorporation assays of *E2F1^{-/-}E2F2^{-/-}E2F3^{+/+}* MEFs that were infected with a control (–) or a Cre retrovirus (+). Serum-starved cells were infected with adenoviruses expressing cyclin E and Cdk2 (E) or with a control adenovirus (C). Adenovirus-infected cells were incubated with BrdU in either 0.2% serum (–) or 15% serum (+) for 24 h, and processed for immunohistochemistry. **e**, *E2F1^{-/-}E2F2^{-/-}E2F3^{+/+}* MEFs were treated as in **a**, and were assayed for E2F DNA-binding activity by EMSA using an E2F-specific ³²P-labelled DNA probe.

of cellular proliferation, the relative importance of E2F-mediated transcriptional repression versus activation remains to be determined. The mechanism by which E2F family members regulate p21^{Cip1} expression and its *in vivo* consequences will be an important topic of future experimentation.

Although a role for E2F in the regulation of the cell cycle and proliferation has been speculated for over a decade^{1,2,26–28}, the complexity of the E2F network has precluded previous loss-of-function mouse models from demonstrating the essentiality for individual E2F family members in mammalian cell proliferation. The ‘conditional’ gene knockout strategy described here allowed us to circumvent detrimental consequences that arise from the inactivation of multiple E2F family members in mice. By making genetic modifications to the entire E2F activator subclass, as opposed to individual members, we provide direct genetic evidence suggesting that E2F activators are essential for cell cycle progression, proliferation and development. □

Methods

Construction of the targeting vector

We used a triple *loxP* vector system (R. Premont, personal communication) to construct our targeting vector. We used a 1.7-kilobase (kb) *EcoRV*–*SpeI* fragment as the short recombination arm, a 2.3-kb *SpeI* fragment, which includes exon 3, as the floxed fragment,

and a 6-kb *SpeI* fragment spanning exons 4 and 5 as the long recombination arm. The final targeting vector shown in Fig. 1a was constructed using standard subcloning procedures. We verified all of the constructs by restriction analysis and sequencing.

Generation of E2F3 knockout mice

Strain 129 E14 embryonic stem (ES) cells were electroporated with 50 μg *NotI* linearized targeting vector. Homologous recombination was selected with G418 and diphtheria toxin and verified by Southern blot and PCR analysis. The properly integrated ES cell clones were transiently transfected with a Cre-expressing plasmid. After negative selection with gancyclovir, clones with either a conventional null allele or a conditional null allele were screened and confirmed by Southern blot and PCR analysis. Appropriate ES cell clones were injected into E3.5 C57BL/6 blastocysts that were subsequently implanted into foster mothers. We bred the resulting chimaeras with C57BL/6 females to achieve germline transmission. Genotypic analysis of offspring was performed by Southern analysis and multiplex PCR using the three primers shown in Fig. 1a (primer A, 5′-GTGGCTGG AAGGGTGCACAG-3′; primer B, 5′-TGAATCATGGACAGAGCCAGG-3′; and primer C, 5′-GATTGATTCTGGGTTGTCAGG-3′).

MEF generation and establishment of MEF cell lines

Primary MEFs were isolated from E13.5 embryos using standard methods. The relative contributions from 129/Sv and C57BL/6 genetic backgrounds in each of the E2F-deficient MEF lines used in this study are very similar. Moreover, the experiments presented here were repeated in at least two different MEF preparations, most of which were from different litters. We generated immortalized cell lines from primary MEFs using the 3T9 cell line preparation protocol.

Retroviral infections

Full-length complementary DNAs for Cre recombinase, Myc-tagged versions of mouse E2F3a and E2F3b and HA-tagged versions of human E2F1, E2F2 and E2F3a were subcloned into the pBabe retroviral vector containing either a puromycin (pBpuro), hygromycin (pBhygro), or phleomycin-resistant gene (pBbleo). High-titre viruses were produced by transient transfection of retroviral constructs into the Phoenix-Eco packaging cell line as described previously²⁹. We infected MEFs with the retrovirus using standard methods. Infected cells were then selected for a total of five days in the presence of one or more antibiotics using the following concentrations: 2.5 μg ml^{−1} for puromycin, 400 μg ml^{−1} for hygromycin and 25 μg ml^{−1} for phleomycin.

Proliferation assays

Immortalized MEFs were seeded at 7 × 10⁴ cells per 60-mm dish. Duplicate plates of cells were counted daily and were replated every 72 h at the same density of the initial plating. Colony formation assays were performed by plating 600 or 3,000 immortalized cells per 100-mm dish with five dishes at each concentration. Cells were cultured in DMEM with 15% fetal bovine serum (FBS) until colonies formed, and then stained with 5 mg ml^{−1} crystal violet in 20% methanol. The number of colonies arising in five plates was determined and the mean and standard deviations of five cultures from one representative experiment are reported. Single colonies were isolated from parallel, 96-well culture plates. DNA was collected from individual colonies and was genotyped by PCR.

Serum starvation and serum stimulation

Subconfluent MEFs were synchronized by incubation in DMEM with 0.2% FBS for either 60 h (for primary MEFs) or 72 h (for immortalized MEFs). Synchronized cells were then stimulated to proliferate by the addition of DMEM supplemented with 15% FBS. Cells were collected at different time points after serum stimulation and were processed for BrdU incorporation assays¹⁴, flow cytometry¹⁴, histone H1 kinase assays¹⁴, EMSA³⁰, western blots, or northern blots. For BrdU incorporation assays, we counted at least 300 4,6-diamidino-2-phenylindole (DAPI) counter-stained nuclei for each time point.

Western and northern blot analyses

Cell protein lysates were separated in SDS acrylamide gels and blotted into polyvinylidene fluoride membranes. Blots were incubated in blotting buffer (5% skim milk in Tris-buffered saline) with antibodies specific for E2F3 (SC-878, Santa Cruz), phospho-Akt (Cell Signalling, no. 9271L), phospho-Erk (Cell Signalling, no. 9101S), p21^{Cip1} (M-19 and C-19, Santa Cruz), p27^{Kip1} (Transduction Laboratory), p53 (14461C, PharMingen and Ab-1, Oncogene), Rb (G3-245, PharMingen), c-Myc-epitope (C-33, Santa Cruz), HA epitope (Y-11, Santa Cruz), cyclin B1 (GNS1, Santa Cruz) and Cdk2 (M2, Santa Cruz). The primary antibodies were then detected using horseradish-peroxidase-conjugated secondary antibodies and ECL reagent as described by the manufacturer (Amersham).

Total RNA for northern blot analysis was isolated using TRIzol (GibcoBRL), and mRNA was subsequently purified using the PolyAtract mRNA isolation system as described by the manufacturer (Promega). Purified mRNA was separated on a 1% agarose gel containing 6% formaldehyde and transferred onto a GeneScreen membrane (NEN Life Science Products). The cDNA probes were labelled using Prime-It RmT (Stratagene) with 50 μCi [α-³²P]dCTP.

Statistical analysis

We performed statistical analyses using Splus 4.5 (Mathsoft). Observed versus expected progeny proportions were tested as deviations from the null binomial distributions conditioned on the fixed number of examined progeny. Confidence intervals were based on standard normal approximations. Cell growth (control versus Cre) kinetics for each

genotype group were compared as follows: the ratios of cell numbers at each time point and the previous time point were compared in control versus Cre to account for serial correlation. Graphical summaries revealed that the ratios for control versus Cre were positively correlated, and we used two-sided *t*-comparisons (9 degrees of freedom) to compare the ratios. The average control versus Cre growth rates showed a monotonically increasing trend across the genotypes shown. We applied a simple nonparametric test based on all possible orderings. Similar techniques were applied for per cent BrdU incorporation after serum stimulation, using the differences of successive observations to account for serial correlation. Percentages were converted to the logarithmic scale and time points after the apparent detection threshold at 8 h were used.

Received 30 July; accepted 25 September 2001.

- Dyson, N. The regulation of E2F by pRB-family proteins. *Genes Dev.* **12**, 2245–2262 (1998).
- Nevins, J. R. Toward an understanding of the functional complexity of the E2F and retinoblastoma families. *Cell Growth Differ.* **9**, 585–593 (1998).
- DeGregori, J., Leone, G., Miron, A., Jakoi, L. & Nevins, J. R. Distinct roles for E2F proteins in cell growth control and apoptosis. *Proc. Natl Acad. Sci. USA* **94**, 7245–7250 (1997).
- Duronio, R. J., O'Farrell, P. H., Xie, J. E., Brook, A. & Dyson, N. The transcription factor E2F is required for S phase during *Drosophila* embryogenesis. *Genes Dev.* **9**, 1445–1455 (1995).
- Royzman, L., Whittaker, A. J. & Orr-Weaver, T. L. Mutations in *Drosophila* DP and E2F distinguish G1–S progression from an associated transcriptional program. *Genes Dev.* **11**, 1999–2011 (1997).
- Field, S. J. *et al.* E2F-1 functions in mice to promote apoptosis and suppress proliferation. *Cell* **85**, 549–561 (1996).
- Humbert, P. O. *et al.* E2F4 is essential for normal erythrocyte maturation and neonatal viability. *Mol. Cell* **6**, 281–291 (2000).
- Humbert, P. O. *et al.* E2F3 is critical for normal cellular proliferation. *Genes Dev.* **14**, 690–703 (2000).
- Leone, G. *et al.* Myc requires distinct E2F activities to induce S phase and apoptosis. *Mol. Cell* **8**, 105–113 (2001).
- Lindeman, G. J. *et al.* A specific, nonproliferative role for E2F-5 in choroid plexus function revealed by gene targeting. *Genes Dev.* **12**, 1092–1098 (1998).
- Rempel, R. E. *et al.* Loss of E2F4 activity leads to abnormal development of multiple cellular lineages. *Mol. Cell* **6**, 293–306 (2000).
- Yamasaki, L. *et al.* Tumor induction and tissue atrophy in mice lacking E2F-1. *Cell* **85**, 537–548 (1996).
- Lees, J. A. *et al.* The retinoblastoma protein binds to a family of E2F transcription factors. *Mol. Cell Biol.* **13**, 7813–7825 (1993).
- Leone, G. *et al.* E2F3 activity is regulated during the cell cycle and is required for the induction of S phase. *Genes Dev.* **12**, 2120–2130 (1998).
- Leone, G. *et al.* Identification of a novel E2F3 product suggests a mechanism for determining specificity of repression by Rb proteins. *Mol. Cell Biol.* **20**, 3626–3632 (2000).
- Lukas, C. *et al.* Accumulation of cyclin B1 requires E2F and cyclin-A-dependent rearrangement of the anaphase-promoting complex. *Nature* **401**, 815–881 (1999).
- Ishida, S. *et al.* Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. *Mol. Cell Biol.* **21**, 4684–4699 (2001).
- Muller, H. *et al.* E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes Dev.* **15**, 267–285 (2001).
- Brehm, A. *et al.* Retinoblastoma protein recruits histone deacetylase to repress transcription. *Nature* **391**, 597–601 (1998).
- Harbour, J. W. & Dean, D. C. Chromatin remodeling and Rb activity. *Curr. Opin. Cell Biol.* **12**, 685–689 (2000).
- Luo, R. X., Postigo, A. A. & Dean, D. C. Rb interacts with histone deacetylase to repress transcription. *Cell* **92**, 463–473 (1998).
- Magnaghi-Jaulin, L. *et al.* Retinoblastoma protein represses transcription by recruiting a histone deacetylase. *Nature* **391**, 601–605 (1998).
- Weintraub, S. J., Prater, C. A. & Dean, D. C. Retinoblastoma protein switches the E2F site from positive to negative element. *Nature* **358**, 259–261 (1992).
- Weintraub, S. J. *et al.* Mechanism of active transcriptional repression by the retinoblastoma protein. *Nature* **375**, 812–815 (1995).
- Zhang, H. S., Postigo, A. A. & Dean, D. C. Active transcriptional repression by the Rb–E2F complex mediates G1 arrest triggered by p16INK4a, TGF β , and contact inhibition. *Cell* **97**, 53–61 (1999).
- Tsai, K. Y. *et al.* Mutation of E2F-1 suppresses apoptosis and inappropriate S phase entry and extends survival of Rb-deficient mouse embryos. *Mol. Cell* **2**, 293–304 (1998).
- Yamasaki, L. *et al.* Loss of E2F-1 reduces tumorigenesis and extends the lifespan of Rb1(+/-) mice. *Nature Genet.* **18**, 360–364 (1998).
- Ziebold, U., Reza, T., Caron, A. & Lees, J. A. E2F3 contributes both the inappropriate proliferation and to the apoptosis arising in Rb mutant embryos. *Genes Dev.* **15**, 386–391 (2001).
- Pear, W. S., Nolan, G. P., Scott, M. L. & Baltimore, D. Production of high-titer helper-free retroviruses by transient transfection. *Proc. Natl Acad. Sci. USA* **90**, 8392–8396 (1993).
- Nevins, J. R., DeGregori, J., Jakoi, L. & Leone, G. in *Methods in Enzymology* (ed. Dunphy, W. G.) 678 (Academic, San Diego, 1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We are grateful to R. Premont for providing the *loxP* vectors. We thank C. Bock for assistance in generating the *E2F3* chimeric mice; C. Brown and L. Jakoi for various technical assistance; and M. Weinstein for critical comments on the manuscript. This work was supported by grants from the National Institutes of Health (NIH) (G.L.). L.W. was supported by an NIH award, C.T. was supported by the Up on the Roof Human Cancer Genetics Postdoctoral Fellowship, J.R.N. is an investigator of the Howard Hughes Medical Institute, and G.L. is a V-Foundation and Pew Charitable Trust Scholar.

Correspondence and requests for materials should be addressed to G.L. (e-mail: Leone-1@medctr.osu.edu).

A phosphate transporter expressed in arbuscule-containing cells in potato

Christine Rausch^{††}, Pierre Daram^{††}, Silvia Brunner[†], Jan Jansa[†], Maryse Laloi[‡], Georg Leggewie[§], Nikolaus Amrhein^{||} & Marcel Bucher[†]

[†] Federal Institute of Technology (ETH) Zurich, Institute of Plant Sciences, Experimental Station Eschikon 33, 8315 Lindau, Switzerland

[§] Max Planck Institute of Plant Molecular Physiology, 14424 Potsdam, Germany

^{||} Federal Institute of Technology (ETH) Zurich, Institute of Plant Sciences, Universitätsstrasse 2, 8092 Zurich, Switzerland

* These authors contributed equally to the work

[‡] Present address: Nestlé Research Center, Lausanne, PO Box 44, Vers-chez-les-Blanc, 1000 Lausanne 26, Switzerland

Arbuscular mycorrhizas are the most common non-pathogenic symbioses in the roots of plants. It is generally assumed that this symbiosis facilitated the colonization of land by plants¹. In arbuscular mycorrhizas, fungal hyphae often extend between the root cells and tuft-like branched structures (arbuscules) form within the cell lumina that act as the functional interface for nutrient exchange. In the mutualistic arbuscular-mycorrhizal symbiosis the host plant derives mainly phosphorus from the fungus, which in turn benefits from plant-based glucose². The molecular basis of the establishment and functioning of the arbuscular-mycorrhizal symbiosis is largely not understood. Here we identify the phosphate transporter gene *StPT3* in potato (*Solanum tuberosum*). Functionality of the encoded protein was confirmed by yeast complementation. RNA localization and reporter gene expression indicated expression of *StPT3* in root sectors where mycorrhizal structures are formed. A sequence motif in the *StPT3* promoter is similar to transposon-like elements, suggesting that the mutualistic symbiosis evolved by genetic rearrangements in the *StPT3* promoter.

Phosphorus is usually taken up in the form of orthophosphate (Pi). It is well known that Pi transport into the root is mediated by a secondary transport mechanism dependent on the activity of the

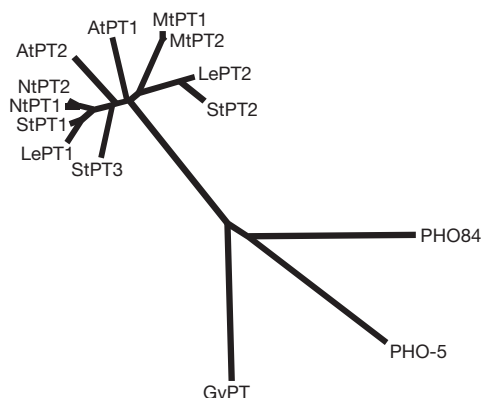


Figure 1 Phylogenetic relationship of *StPT3* with other Pi transporters. Phylogenetic analysis of plant Pht1 and related high-affinity Pi transporters from fungal species. Protein names are followed by SwissProt (SP) or GenBank (GB) accession numbers:

LePT1 (O24029, SP) and LePT2 (O22549, SP) from *Lycopersicon esculentum*; NtPT1 (AAF74025, GB) and NtPT2 (BAA86070, GB) from *Nicotiana tabacum*; StPT1 (Q43650, SP), StPT2 (Q41479, SP) and StPT3 (AJ318822, GB) from *Solanum tuberosum*; MtPT1 (O22301, SP) and MtPT2 (O22302, SP) from *Medicago truncatula*; AtPT1 (Q96302, SP) and AtPT2 (Q96303, SP) from *Arabidopsis thaliana*; PHO84 (P25297, SP) from *Saccharomyces cerevisiae*; PHO-5 (L36127, GB) from *Neurospora crassa*, and GvPT (U38650, GB) from *Glomus versiforme*.

plasma membrane H^+ -ATPase. The H^+ -ATPase protein is localized in the plant membrane around arbuscule hyphae, which corroborates the existence of nutrient transport activities at this root–fungus interface³. Plant Pi transporters have recently been cloned and shown to be expressed in the root epidermis (the rhizodermis, including root hairs) of tomato and *Medicago truncatula* where they are involved in Pi uptake at the root–soil interface^{4,5}. The encoded proteins are orthologous to GvPT that is expressed in extraradical hyphae of the arbuscular-mycorrhizal fungus *Glomus versiforme*⁶ and were assigned to the Pht1 family of Pi transporters in plants⁷. Even though steady-state messenger RNA levels of Pht1 transporters are generally lower in roots of mycorrhizal than in non-mycorrhizal plants^{8,9}, mRNA of the Pi transporter LePT1 was detected in arbuscule-containing cells of mycorrhizal tomato⁸, suggesting altered cellular localization of Pht1 transporters during fungal colonization in mycorrhizal roots.

A screen for potato (*Solanum tuberosum* L. var. Désirée) Pi transporter genes previously led to the isolation of two complete complementary DNAs, *StPT1* and *StPT2*¹⁰, and a partial cDNA clone, *StPT3* (G.L., unpublished work). Here, we have investigated the function and cell-specific expression of the *StPT3* gene to test the hypothesis that the *StPT3* protein functions as an arbuscular-mycorrhizal symbiosis-specific Pi transporter.

High-stringency genomic DNA gel blot analysis using the partial *StPT3* cDNA as a probe indicated the presence of *StPT3* as a single or low-copy gene in the haploid potato genome (data not shown). A screen of a potato genomic DNA library resulted in the cloning of a 3,404 base-pair (bp) DNA fragment consisting of an intron-less

StPT3 sequence encoding a 535-amino-acid protein. The open reading frame is flanked by 247 bp of untranslated sequence at the 3' region based on sequence information from the partial *StPT3* cDNA clone. The 5' flanking region encompasses 1,727 bp containing a 9-bp dyad-symmetric sequence (CAATTATTG) for *in vitro* binding of the Athb-1 homeodomain-leucine zipper (HD-Zip-1) domain at position –1,227 and a first predicted TATA box 177 bp upstream of the initiator ATG (Fig. A of the Supplementary Information).

An unrooted phylogenetic tree diagram demonstrates the close relationship between the *StPT3* protein and members of the plant Pht1 family and high-affinity Pi transporters of fungal origin (Fig. 1). Like its other family members, *StPT3* contains 12 putative membrane-spanning domains, hydrophilic amino and carboxy termini, and a hydrophilic loop between transmembrane segments six and seven. *StPT3* is similar to other Pht1 sequences and shares a putative glycosylation site in the predicted sixth transmembrane domain and two intracellular phosphorylation sites with all polypeptides compared in Fig. 1 (Figs A and B of the Supplementary Information). The presence of these motifs suggests post-translational modification of the protein likely to be involved in structure formation and/or the regulation of transport activity. To obtain biochemical evidence for the function of *StPT3*, the yeast mutant PAM2, which is defective in the two high-affinity Pi transporter genes *PHO84* and *PHO89* (ref. 11), was used for complementation analysis (Fig. 2). In uptake experiments with ³²Pi, the rate of net transport was linear with time during the first 5 min of uptake under the conditions applied (data not shown) and was significantly higher in cells expressing *StPT3* or the tomato Pi transporter cDNA *LePT1* than in mutant cells (Fig. 2a). At 20 μ M Pi, the mutant cells expressing *StPT3* were able to grow, whereas no growth was observed in the case of the cells carrying the empty expression vector (Fig. 2b). The cells grew well at 2 mM Pi concentrations, probably owing to the activity of the endogenous yeast low-affinity Pi uptake system. Uptake velocities measured at different Pi concentrations indicated that Pi uptake mediated by *StPT3* followed Michaelis–Menten kinetics (Fig. 2c), exhibiting an apparent K_m value of $64 \pm 3 \mu$ M ($n = 3$). We are aware that the apparent K_m obtained from expression in yeast need not necessarily reflect the K_m in living plants. In accordance with previous work, we conclude that *StPT3*, similar to the tomato Pi transporter LePT1 and presumably other Pht1 proteins⁷, functions as a plasma membrane high-affinity Pi transporter in yeast cells.

We could not detect *StPT3* transcripts in non-mycorrhizal plants.

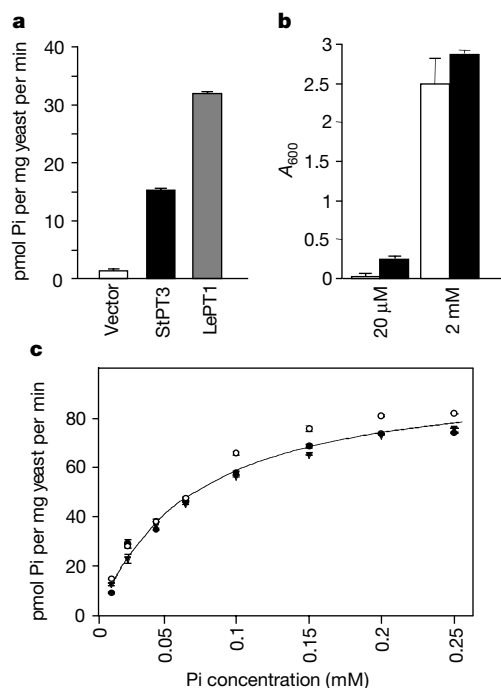


Figure 2 Complementation of Pi uptake-deficient yeast and kinetics of Pi uptake. **a**, Pi uptake into yeast PAM2 mutant cells transformed with the expression vector harbouring either no insert (vector, white column), *StPT3* cDNA (StPT3, black column), or *LePT1* cDNA (LePT1, grey column). **b**, Growth of *StPT3*-expressing and PAM2 mutant yeast cells as reflected by absorbance (A) of cultures grown in the presence of either 20 μ M or 2 mM Pi, respectively. Means and standard errors of means of three replicate determinations consisting of three measurements each are shown in **a** and **b**. **c**, Nonlinear regression of Pi uptake of strain StPT3 versus external Pi concentration at pH 5 that was used to estimate the K_m value. Symbols designate mean values from three replicates of three measurements each.

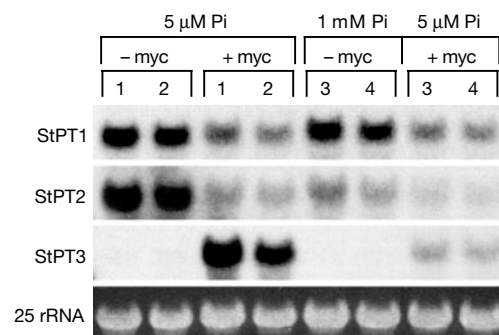


Figure 3 *StPT3* transcript abundance in mycorrhizas. Pi transporter transcript levels in roots cultured in a split-root system. Randomly labelled cDNA probes, as indicated at left, hybridized to RNA on the blot from parts of roots of two individual plants per split-root system (1, 2 and 3, 4) that were cultured without (–myc) or with (+myc) *G. intraradices*, respectively. Both compartments of one system were irrigated with 5 μ M Pi, but the non-inoculated (–myc) compartment of the second system was irrigated with 1 mM Pi. 25S rRNA served as a control for equal loading of the gel and RNA integrity.

Table 1 Percentage colonization in the split-root system

Pi (μM) n-i/i	Colonization of the roots with specific arbuscular-mycorrhizal structures (%)			
	Arbuscules	Vesicles	Hyphae	Total
1,000/5	20.0 $\pm$ 3.3	3.3 $\pm$ 2.9	55.6 $\pm$ 9.7	76.2 $\pm$ 7.8
5/5	57.8 $\pm$ 4.4	27.1 $\pm$ 4.4	23.5 $\pm$ 2.4	87.8 $\pm$ 5.0

Fungal colonization of the root as a function of the Pi concentration in the nutrient medium of the non-inoculated (n-i) and inoculated (i) root compartment is expressed as the percentage of the root length containing various mycorrhizal structures determined by a grid intersect method²⁹. For further explanation, see text.

Therefore, *StPT3* expression was evaluated in split-roots of either Pi-deprived or Pi-supplied potato plants. One part of the root was colonized by the arbuscular-mycorrhizal fungus *Glomus intraradices* at low external Pi concentration, while the other part was not inoculated with the fungus and was cultivated at either low or high Pi concentrations, respectively. High-stringency RNA gel blot analysis revealed the presence of a single transcript of roughly 2,000 nucleotides in mycorrhizal parts of roots which was undetectable in either leaves (data not shown) or non-mycorrhizal parts of the same roots (Fig. 3). As would be consistent with fungal infection, the abundance of the transcripts of the Pi transporter genes *StPT1* and *StPT2* was reduced, probably as a result of the plant-fungus interaction and/or an improved Pi status of the mycorrhizal roots. Moreover, while *StPT1* mRNA levels remained high in non-mycorrhizal roots irrespective of the Pi concentration, *StPT2* expression was reduced at high Pi conditions. With increasing Pi

concentration in the non-mycorrhizal compartment, *StPT3* mRNA abundance decreased in mycorrhizal roots correlating with reduced arbuscule and vesicle formation and increased presence of fungal hyphae (Table 1 and Fig. 3). These data led us to conclude that the three *Pht1* genes of the potato are differentially regulated and that the *StPT3* gene is locally induced on colonization of the root with arbuscular-mycorrhizal fungi. Moreover, *StPT3* steady-state mRNA concentrations correlate with arbuscule and vesicle formation, although vesicles are not likely to be sites of Pi transfer.

Next, we analysed the cellular localization of *StPT3* mRNA in the potato mycorrhiza using non-radioactive *in situ* hybridization with a digoxigenin-labelled antisense RNA probe. High-stringency hybridization yielded specific signals with the antisense probe, but no labelling with the sense probe (Fig. 4a–c). *StPT3* mRNA labelling was most prominent in arbuscule-containing cells (Fig. 4a, b), so the protein translated from the *StPT3* mRNA is presumably located in the periarbuscular membrane, which is of plant origin, and is involved in the ultimate absorption of phosphorus by the plant. Ribosomal RNA was detectable in all cells (data not shown). A qualitative examination of mycorrhiza segments along the root indicated that *StPT3* mRNA is present at varying abundance in different root sectors (data not shown), correlating with the short lifespan of a single arbuscule of a few days². To support these data and to characterize the sequences required for this highly specific expression pattern, we fused the upstream flanking DNA sequence of the *StPT3* genomic clone with an exact fusion at the initiator ATG to the β -glucuronidase (*GUS*) reporter gene¹². This chimaeric gene

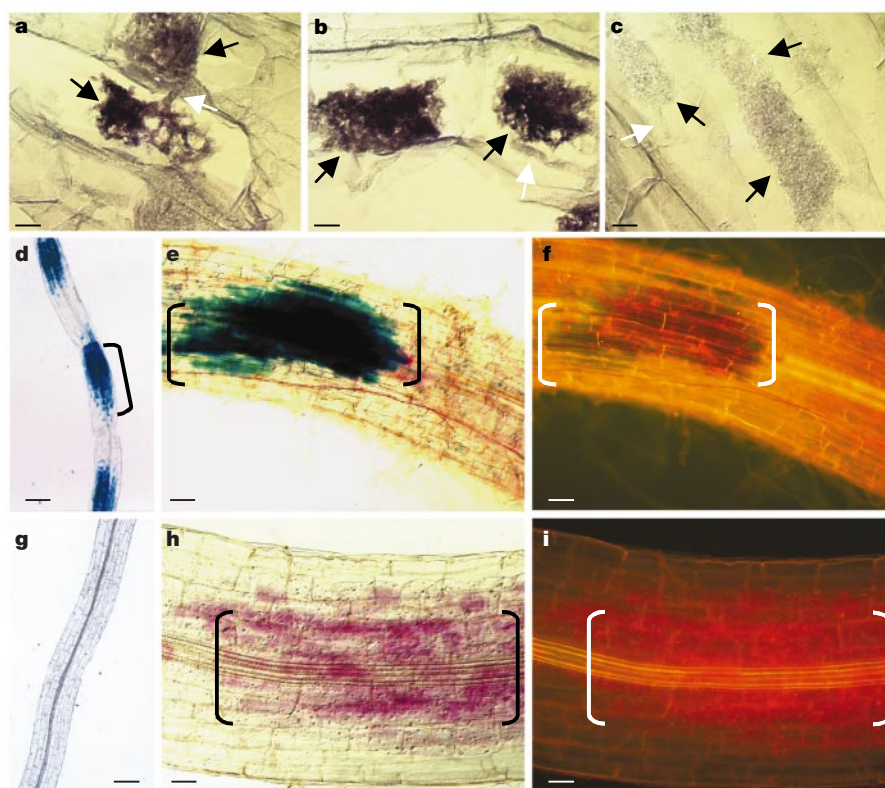


Figure 4 Localization of *StPT3* transcripts and reporter gene activity in arbuscule-containing root sectors. **a–c**, *In situ* hybridization showing root sections hybridized with the *StPT3* antisense (**a**, **b**) or sense (**c**) probe, respectively. The purple dye in **a** and **b** indicates *StPT3* mRNA. Black arrows indicate arbuscules, white arrows indicate hyphae. **d**, **e**, Indigo blue GUS staining in mycorrhizal roots carrying the *StPT3/GUS* chimaeric gene. **f**, Dark-blue GUS co-localizes with fungal colonization, reflected by red fluorescence of fungal structures stained with acid fuchsin. Brackets indicate stained

sectors. **g–i**, Absence of GUS expression in mycorrhizal control roots carrying a minimal promoter/*GUS* chimaeric gene and stained for GUS activity. Brackets indicate purple-stained (**h**) or red-fluorescent (**i**) sectors colonized with *G. intraradices*. Sections and roots were analysed by bright-field microscopy except for **f** and **i**, where fluorescence under ultraviolet illumination was detected. Identical root segments were photographed in the **e**, **f** and **h**, **i** pairs, respectively. Scale bars: **a–c**, 10 μm ; **d**, **g**, 200 μm ; **e**, **f**, **h**, **i**, 40 μm .

was introduced into potato via *Agrobacterium tumefaciens*-mediated gene transfer. Five transgenic lines were subsequently inoculated with *G. intraradices*. Histochemical staining for GUS activity in mycorrhizal roots indicated that the stain was detectable in distinct areas, (Fig. 4d–f) which corresponds to the zonation of fungal root colonization, whereas non-colonized neighbouring cells, or mycorrhizal roots transformed with a minimal promoter/-GUS chimaeric gene (Fig. 4g–i) never stained blue. Co-localization of GUS staining with arbuscules or hyphae was demonstrated using acid fuchsin (Fig. 4f) and chlorazol black staining (data not shown) of fungal structures. Thus, *StPT3* gene expression is most likely arbuscular-mycorrhizal colonization-specific, and is associated with root sectors where arbuscules are formed.

A database search was carried out to look for motifs shared by the *StPT3* promoter with promoters of other plant genes. A Blast search (<http://www.ncbi.nlm.nih.gov/blast/>) revealed an element of 146-nucleotides sharing high similarity with a sequence motif present in disease resistance loci from plants (Fig. C of the Supplementary Information). The 146-nucleotide element is contained in a longer genomic fragment of 431 bp composed of terminal inverted repeats and subterminal repeats that shares analogies with miniature inverted-repeat transposable elements¹³. Homology is most obvious to loci conferring resistance to bacteria in *Lycopersicon pimpinellifolium* and *L. esculentum* cv. VFNT Cherry (Pto), and to viral (*rx* gene), nematode (*Gpa2* gene, *Mi-copy2* gene homologue) and aphid (*Mi-copy2* gene homologue) resistance genes, respectively, from potato and tomato. Owing to the observed homologies, we designated the conserved element MRR (mycorrhiza and resistance-related element). Two highly conserved subdomains within the MRR element, MRR1 and MRR2, localize, either individually or in combination, in varying polarity to inverted repeat elements in the promoter region within generally 2,000 bases upstream of the resistance gene's coding region (data not shown). These results support the hypothesis that the MRR element has inserted into the *StPT3* promoter during evolution.

StPT3 belongs to a gene family of ancient evolutionary origin and exhibits high homology to *G. versiforme* GvPT and plant Pht1 transporters: GvPT is involved in Pi uptake into extraradical fungal hyphae and Pht1 members are responsible for Pi transport at the root–soil interface into epidermal cells, including root hairs^{4,5}. The present work expands this model and indicates that proteins orthologous to *StPT3* mediate Pi acquisition at three interfaces of a mycorrhiza, that is, at the extraradical hyphae, the root epidermis, and cells containing arbuscules or hyphae. We speculate that the gain of function of a host-plant high-affinity Pi transporter at the root–fungal interface had an important role in the evolution of the mutualistic arbuscular-mycorrhizal symbiosis from a saprophytic interaction and thus in the development of vascular land plants. There is evidence indicating that ectomycorrhizal symbionts have evolved repeatedly from saprotrophic precursors¹⁴. We propose that rearrangements in the promoter sequence of an ancient Pi transporter gene in a dynamic genome caused mycorrhiza-specific gene expression, probably as a consequence of the insertion of a transposable element (such as those containing a MRR element)¹⁵.

Arbuscular-mycorrhizal fungi are an important and abundant component of soil biota in most terrestrial ecosystems. This group of organisms also represents a distinct trophic group (biotrophic) from most of the rest of the soil microbial biomass (saprophytic, parasitic or pathogenic). Biotrophs establish long-term feeding relationships (mutualistic or parasitic) with the living cells of their hosts, rather than killing the host cells as part of the infection process. Arbuscular-mycorrhizal fungi can influence plant community composition by differentially affecting the growth of different plant species^{16,17}, and mycorrhizal colonization may influence plant community diversity¹⁸. Moreover, it was recently emphasized that arbuscular-mycorrhizal fungal diversity is an important contributor

to plant biodiversity, ecosystem variability and productivity¹⁹. Thus, from an ecological point of view, *StPT3* orthologues are likely to be a principal molecular determinant of establishment, function and stability of terrestrial ecosystems. The *StPT3* gene can now be evaluated as a molecular marker for investigating the functional state of the arbuscular-mycorrhizal symbiosis. The study of *StPT3* gene regulation may provide a clue to the mechanisms of communication between fungus and host plant. Moreover, *StPT3* orthologues from crop plants may be used as molecular markers for the effective use of endomycorrhizas in plant production systems. □

Methods

DNA cloning

A partial *StPT3* cDNA was cloned as described¹⁰. The cDNA was used to screen a genomic potato library (Clontech). λDNA was prepared as described²⁰. A genomic DNA fragment of about 3.4 kilobases was isolated and cloned into pBluescript SK[−] (Stratagene). Sequencing of both strands was performed using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer Applied Biosystems) according to the manufacturer's protocol and revealed a 900-bp overlap with the partial *StPT3* cDNA sequence including the full open reading frame and an extension into the 5' upstream region of the *StPT3* gene. The open reading frame as derived from the genomic sequence, including the 5' untranslated region, was confirmed via double-stranded sequencing of the partial *StPT3* cDNA and four cDNAs generated independently by polymerase chain reaction (PCR) after reverse transcription of RNA (RT–PCR) (1 cDNA) or 5' rapid amplification of cDNA ends (RACE; 3 cDNAs). An *NcoI* restriction site was introduced at the initiator ATG of the *StPT3* coding region via PCR according to standard procedures. The promoter was then sequenced and cloned with an exact fusion at the initiator ATG to the β-glucuronidase (*GUS*) marker gene/nopaline synthase (*nos*) terminator^{12,21} from plasmid 5'AKT1-320.X²² into the binary vector Bin19 (ref. 23). The *StPT3* coding region was subcloned into the *SmaI*–*BamHI* sites of the yeast shuttle vector 181A1NE³. *LePT1*, encoding a tomato high-affinity H⁺/Pi cotransporter, was used as a control.

Yeast manipulation

Competent yeast cells were prepared and transformed as described²⁴. Growth experiments, ³²Pi uptake studies, and calculation of the apparent *K_m* were performed as described²⁵.

Plant transformation and growth conditions

The *StPT3* promoter–*GUS* chimeric gene was introduced in *Agrobacterium tumefaciens* strain C58C1 containing the pGV2260 plasmid²⁶ via electroporation. Transformation of potato (*Solanum tuberosum* var. Désirée) was performed using *A. tumefaciens*-directed gene transfer basically as described²⁷. Wild-type or transgenic potato plants from tissue culture were first adapted to greenhouse conditions during one week in sterilized quartz sand and then transferred to a soil/quartz sand mixture (1:10) containing either isolated spores (~1,000 spores per inoculation) or maize roots (see also <http://www.bio.ukc.ac.uk/beg/>) infected with *G. intraradices* Schenck and Smith (BEG75). Plants were fertilized with half-strength Hoagland medium²⁸ containing 1 mM NH₄H₂PO₄ before inoculation, and either 1 mM or 5 μM NH₄H₂PO₄, respectively, during arbuscular-mycorrhizal fungal colonization for six to eight weeks, using drop irrigation, in two independent experiments. In split-root studies, the root system of potato plants was separated into two equal parts and the parts were individually cultivated in two adjacent compartments, with infected maize root segments added to one of the compartments. Percentage of colonization was calculated with a gridline intersect method²⁹ using trypan blue/methylene blue (both 0.05%, w/v) stained mycorrhiza after clearing in KOH (see below) and bright field microscopy at 400-fold magnification.

RNA and reporter gene expression analysis

RNA extraction and RNA gel blot analysis were performed as described previously⁵. *StPT1*, *StPT2* and *StPT3* partial cDNAs of roughly 730 to 1,000 bp in length and including the 3' non-translated region were used as radioactively labelled probes. Twenty micrograms of total RNA was used from each sample. Hybridization was with 5 × SSC, 5 × Denhardt's, 0.5% SDS (w/v) at 60 °C with a final wash using 0.1 × SSC and 0.1% SDS at 68 °C. Localization of mRNA in 15-μm-thick root sections was performed by *in situ* hybridization essentially as described previously⁵. Digoxigenin-labelled antisense and sense RNAs were generated from the pBluescript SK(–) plasmid (Stratagene) containing the original *StPT3* partial cDNA using the Riboprobe Combination System from Promega. Dye formation using the anti-digoxigenin antibody coupled to alkaline phosphatase occurred at room temperature overnight. Stained sections were analysed by light microscopy (Olympus AZ70, Olympus Optical Schweiz). For GUS assays, root material was incubated in 0.1% X-Gluc (Biosynth) and 0.1% Triton X-100 (Fluka Chemie) in 0.05 M sodium phosphate buffer, at pH 7.2 and 37 °C after vacuum infiltration. The stained material was transferred directly to 10% KOH (w/v) at 90 °C for 1 h to clear the tissue and then, when needed, counterstained for 1 h at 90 °C with either 0.03% chlorazol black²⁹ or with 0.1% acid fuchsin³⁰ dissolved in lactoglycerol (lactic acid:glycerol:water, 1:1:1), then destained in 50% glycerol. All stainings were visualized by light or fluorescent microscopy.

Computational analysis

Multiple sequence alignment and topology prediction were done as described²⁵. Phylo-

genetic tree analysis was performed with the Phylip drawtree program at the Pasteur Institute (<http://bioweb.pasteur.fr/seqanal/interfaces/drawtree.html>).

Received 22 June; accepted 27 September 2001.

1. Remy, W., Taylor, T. N., Hass, H. & Kerp, H. Four-hundred-million-year-old vesicular arbuscular mycorrhizae. *Proc. Natl Acad. Sci. USA* **91**, 11841–11843 (1994).
2. Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* (Academic, San Diego, California, 1997).
3. Gianinazzi-Pearson, V., Arnould, C., Oufattole, M., Arango, M. & Gianinazzi, S. Differential activation of H⁺-ATPase genes by an arbuscular mycorrhizal fungus in root cells of transgenic tobacco. *Planta* **211**, 609–613 (2000).
4. Chiou, T. J., Liu, H. & Harrison, M. J. The spatial expression patterns of a phosphate transporter (MtPT1) from *Medicago truncatula* indicate a role in phosphate transport at the root/soil interface. *Plant J.* **25**, 281–293 (2001).
5. Daram, P., Brunner, S., Persson, B. L., Amrhein, N. & Bucher, M. Functional analysis and cell-specific expression of a phosphate transporter from tomato. *Planta* **206**, 225–233 (1998).
6. Harrison, M. J. & Van Buuren, M. L. A phosphate transporter from the mycorrhizal fungus *Glomus versiforme*. *Nature* **378**, 626–629 (1995).
7. Bucher, M., Rausch, C. & Daram, P. Molecular and biochemical mechanisms of phosphorus uptake into plants. *J. Plant Nutr. Soil Sci.* **164**, 209–217 (2001).
8. Rosewarne, G., Barker, S., Smith, S., Smith, E. & Schachtman, D. A *Lycopersicon esculentum* phosphate transporter (LePT1) involved in phosphorus uptake from a vesicular-arbuscular mycorrhizal fungus. *New Phytol.* **144**, 507–516 (1999).
9. Liu, H., Trieu, A. T., Blaylock, L. A. & Harrison, M. J. Cloning and characterization of two phosphate transporters from *Medicago truncatula* roots: Regulation in response to phosphate and to colonization by arbuscular mycorrhizal (AM) fungi. *Mol. Plant-Microbe Interact.* **11**, 14–22 (1998).
10. Leggewie, G., Willimitter, L. & Riesmeier, J. W. Two cDNAs from potato are able to complement a phosphate uptake-deficient yeast mutant: Identification of phosphate transporters from higher plants. *Plant Cell* **9**, 381–392 (1997).
11. Martinez, P. & Persson, B. L. Identification, cloning and characterization of a derepressible Na⁺-coupled phosphate transporter in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **258**, 628–638 (1998).
12. Jefferson, R. A., Kavanagh, T. A. & Bevan, M. W. GUS fusions: β -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **3**, 3901–3907 (1987).
13. Wessler, S. R., Bureau, T. E. & White, S. E. LTR-retrotransposons and MITEs—Important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* **5**, 814–821 (1995).
14. Hibbett, D. S., Gilbert, L. B. & Donoghue, M. J. Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes. *Nature* **407**, 506–508 (2000).
15. Selinger, D. A. & Chandler, V. L. Major recent and independent changes in levels and patterns of expression have occurred at the *b* gene, a regulatory locus in maize. *Proc. Natl Acad. Sci. USA* **96**, 15007–15012 (1999).
16. Varma, A. in *Endomycorrhiza* (eds Varma, A. & Hock, B.) 521–556 (Springer, Berlin, 1999).
17. Johnson, N. C., Graham, J. H. & Smith, F. A. Functioning of mycorrhizal associations along the mutualism–parasitism continuum. *New Phytol.* **135**, 575–585 (1997).
18. Grime, J. P., Mackey, J. M. L., Hillier, S. H. & Read, D. J. Floristic diversity in a model system using experimental microcosms. *Nature* **328**, 420–422 (1987).
19. van der Heijden, M. et al. Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* **396**, 69–72 (1998).
20. Lockett, T. A bacteriophage lambda DNA purification procedure suitable for the analysis of DNA from either large or multiple small lysates. *Anal. Biochem.* **185**, 230–234 (1990).
21. Depicker, A., Stachel, S., Dhaese, P., Zambryski, P. & Goodman, H. M. Nopaline synthase: transcript mapping and DNA sequence. *J. Mol. Appl. Genet.* **1**, 561–573 (1982).
22. Bariola, P. et al. The *Arabidopsis* ribonuclease gene *RNS1* is tightly controlled in response to phosphate limitation. *Plant J.* **6**, 673–685 (1994).
23. Bevan, M. Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acids Res.* **12**, 8711–8721 (1984).
24. Dohmen, R. J., Strasser, A. W. M., Höner, C. B. & Hollenberg, C. P. An efficient transformation procedure enabling long-term storage of competent yeast cells of various yeast genera. *Yeast* **7**, 691–692 (1991).
25. Daram, P. et al. *PhT2* encodes a low-affinity phosphate transporter from *Arabidopsis*. *Plant Cell* **11**, 2153–2166 (1999).
26. Deblaere, R. et al. Efficient octopine Ti plasmid-derived vectors for *Agrobacterium*-mediated gene transfer to plants. *Nucleic Acids Res.* **13**, 4777–4788 (1985).
27. Rocha-Sosa, M. et al. Both developmental and metabolic signals activate the promoter of a class I patatin gene. *EMBO J.* **8**, 23–29 (1989).
28. Hoagland, D. & Broyer, T. General nature of the process of salt accumulation by roots with description of experimental methods. *Plant Physiol.* **11**, 471–507 (1936).
29. Brundrett, M. C., Piché, Y. & Peterson, R. L. A new method for observing the morphology of vesicular-arbuscular mycorrhizae. *Can. J. Bot.* **62**, 2128–2134 (1984).
30. Merryweather, J. W. & Fitter, A. H. A modified procedure for elucidating the structure of the fungal partner in a vesicular-arbuscular mycorrhiza. *Myc. Res.* **95**, 1435–1437 (1991).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank R. Ruh for support with microscopic analysis; B. Regierer and J. Kossmann for the potato genomic library; and T. Mandel and C. Kuhlmeier for the 35S minimal promoter. This work was supported by the Swiss National Science Foundation, and Hoffmann-La Roche, Switzerland.

Correspondence and requests for materials should be addressed to M.B. (e-mail: marcel.bucher@ipw.biol.ethz.ch). The StPT3 gene and protein sequences have been deposited under GenBank accession number AJ318822.

Crystal structure of the tricorn protease reveals a protein disassembly line

Hans Brandstetter*, Jeong-Sun Kim*, Michael Groll & Robert Huber

Abteilung Strukturforschung, Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany

* These authors contributed equally to this work

The degradation of cytosolic proteins is carried out predominantly by the proteasome, which generates peptides of 7–9 amino acids long¹. These products need further processing. Recently, a proteolytic system was identified in the model organism *Thermoplasma acidophilum* that performs this processing^{2,3}. The hexameric core protein of this modular system, referred to as tricorn protease, is a 720K protease that is able to assemble further into a giant icosahedral capsid, as determined by electron microscopy⁴. Here, we present the crystal structure of the tricorn protease at 2.0 Å resolution. The structure reveals a complex mosaic protein whereby five domains combine to form one of six subunits, which further assemble to form the 3-2-symmetric core protein. The structure shows how the individual domains coordinate the specific steps of substrate processing, including channelling of the substrate to, and the product from, the catalytic site. Moreover, the structure shows how accessory protein components might contribute to an even more complex protein machinery that efficiently collects the tricorn-released products.

The hexameric D3-symmetric tricorn protein is assembled by two perfectly staggered and interdigitating trimeric rings, with each subunit of one ring forming contacts almost exclusively with the two subunits of the other ring related by the molecular diads. The toroid structure has the shape of a distorted hexagon. The dimers forming the long sides of the hexagon are intimately intertwined. By contrast, the interface of the subunits forming the short side is very flat. As a consequence, the two resulting subunit interfaces differ drastically in their respective contact area (6,233 and 1,683 Å²) and assemble the tricorn protein as a trimer of dimers (Fig. 1a).

The positions of all but the terminal amino acids are defined in the electron density (M39–N1061). A single subunit is further divided into five subdomains (Fig. 1b). The amino-terminal part of the protein folds as a six-bladed β propeller (M39–D310, referred to here as β 6) and is followed sequentially by a seven-bladed β propeller (A326–K675, β 7). A PDZ-like domain (R761–D855) is interspersed between the two carboxy-terminal mixed α - β domains (S681–G752 and R856–N1061, denoted C1 and C2). Both β 6 and β 7 are topologically unclosed. Such an open 'Velcro'-like propeller structure has been observed only in the crystal structure of prolyl oligopeptidase (POP)⁵. β 7 carries prominent insertions on blades 4, 5 and 6, which mostly mediate contacts to the diad-related subunit. The propeller structures are oriented roughly antiparallel to each other, thereby embracing the two loosely packed C-terminal domains C1 and C2, and partly the PDZ-like domain (Fig. 1b).

Despite their limited domain interface (455 Å²), the core domains C1 and C2 are closely associated by a small sequential insertion within C2 (G993–P1009), integrated into the C1 domain. A further prominent sequence insertion, I933–P947, is structurally integrated into the diad-related C2 domain (and vice versa, Fig. 1b). It reflects the intimate dimer contact on a structural level. The C1 and C2 domains are of mixed α - β type and are, together with the PDZ domain, structurally related to the D1P protease⁶.

Within the hexameric framework, three β 7 propeller domains are arranged around the molecular three-fold axis, where they line the entrance to the central pore (Fig. 1a). Most of the contacts

determining the relative arrangement of the $\beta 7$ propellers are mediated by the intervening PDZ domains of the diad-related subunits. Each of the three short sides of the hexagon is formed by two $\beta 6$ propellers.

To elucidate the amino acids crucial for its catalytic activity, we cocrystallized tricorn with a series of chloromethyl ketone-based inhibitors, including *N*- α -tosyl-L-lysyl-chloromethyl ketone (TLCK) and *N*-tosyl-L-phenylalanyl chloromethyl ketone (TPCK), for which we confirmed inhibitory efficacy (data not shown). For both inhibitors, we observed continuous electron density connecting to the side chain of S965 that was unambiguously fitted by the respective inhibitor. S965 is positioned at a helix entrance within subdomain C2 (Figs 1b and 2b). The uncapped amide nitrogen of D966 forms, together with that of G918, the oxyanion hole that is occupied by a water molecule in the uninhibited structure. H746 is

positioned ideally to activate the catalytic S965 at a hydrogen-bonding distance of 2.7 Å. However, in none of the inhibitor complexes could we observe a covalent linkage of H746 to the inhibitor. We confirmed both S965 and H746 to be crucial for catalysis by constructing the single-site mutants S965A and H746A, both of which are amidolytically inactive. H746 is correctly oriented by the O γ of S745, which in turn is polarized by E1023, thus establishing a catalytic tetrad (see Supplementary Information Fig. 4a). The arrangement of S965, H746 and the oxyanion hole suggests that steps of peptide bond hydrolysis follow the classical sequence of the trypsin-like serine proteases, namely the formation of the tetrahedral adduct, the acyl-enzyme complex, and hydrolysis.

A quite unexpected feature present in the TLCK tricorn complex structure was a continuous electron density connected to the side chain of D966. As no density was visible in the native (uninhibited)

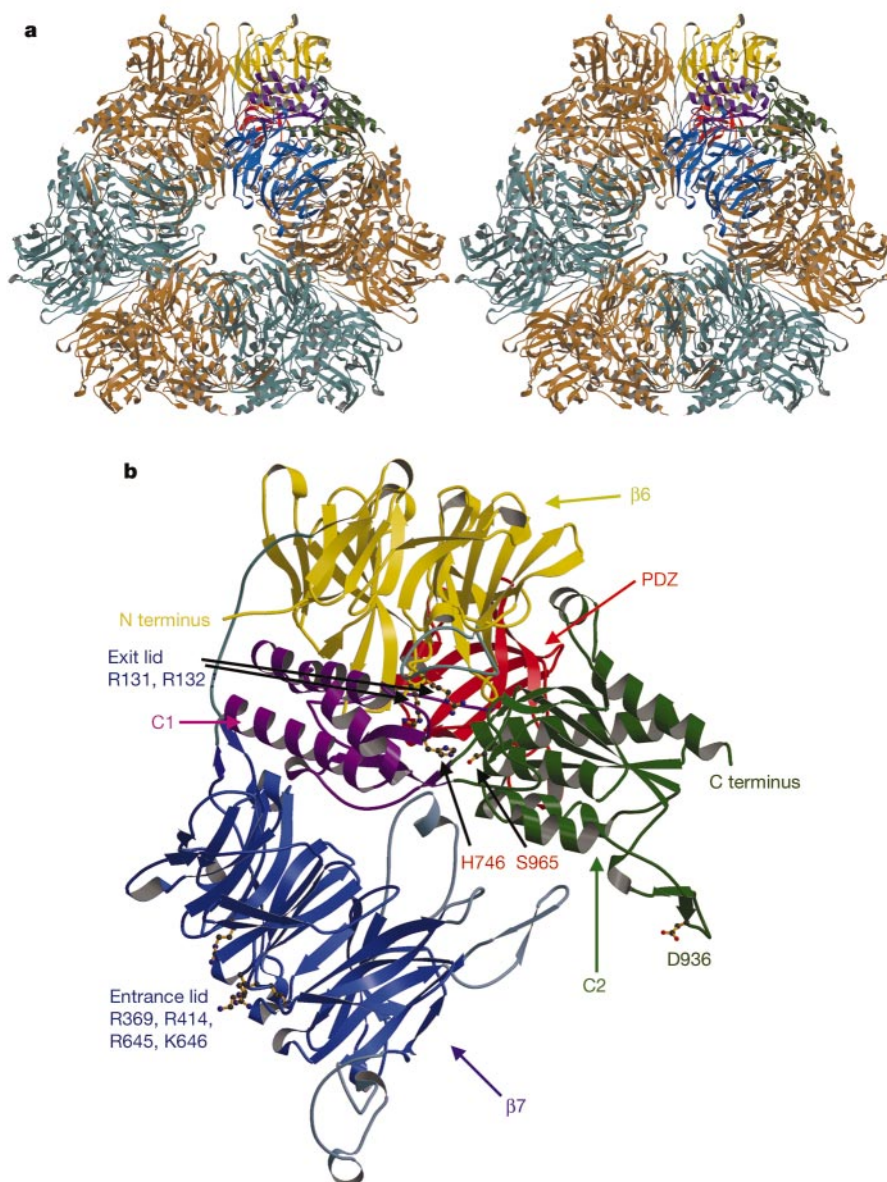


Figure 1 Structure of tricorn protease. **a**, Ribbon representation of tricorn protease viewed along the molecular three-fold axis. Individual subunits are distinguished by colour. The overall dimensions of the molecule are 160 Å within the plane normal to the three-fold axis and 88 Å parallel to it. The conically shaped central pore measures 45 Å in diameter at its entrance and 20 Å close to the centre of the molecule. All figures were prepared with the programs Molscript and Raster3D^{18,19}. **b**, Representation of the

highlighted subunit of **a** using identical colour coding and a similar orientation. The residues forming the propeller lids are displayed as well as the catalytic residues H746 (C1, magenta) and S965 (C2, green). D936, positioned on the C2 domain, confers specificity for basic substrate residues in the active site of the diad-related subunit (compare with **a**).

structure and in a tricorn crystal soaked with lysine, we conclude that the carboxylate side chain of D966 serves as the nucleophile to attack the chloromethyl ketone, a reaction described for other systems⁷.

Tricorn has been shown to exhibit both tryptic and chymotryptic specificities². The X-ray structure reveals that specificity for basic P1 residues (preceding the cleavable bond) is conferred by D936, which is provided by the diad-related subunit (Fig. 2). Here, the previously described structural linkage (trimer of the dimers) is translated into functional cooperativity within the dimers. In the uninhibited 2.0 Å crystal structure, the specificity residue D936 is mobile, however, reflected by an increased crystallographic temperature factor and deviation from noncrystallographic symmetry (NCS). Consistent with this, the side chain of D936 in the TPCK complex structure adopts an alternative rotamer to allow the TPCK phenyl ring to freely access the hydrophobic niche formed by I969, V991 and F1013 (see Supplementary Information Fig. 4b). D936 thus serves as a substrate specificity switch accommodating both hydrophobic and basic P1 residues.

In addition to the described ionic or hydrophobic interactions of the side chains, the P1 residue is held in place by main-chain interactions with the oxyanion hole (G918, D966). P2–P4 residues presumably use unsaturated main-chain hydrogen bonds at the strand I994–P996 (S in Fig. 2b), and further interactions might occur in the $\beta 7$ propeller channel, as described for galactose oxidase⁸. These substrate-recognition sites are rather unrestricted, in accord with tricorn's broad substrate specificity^{2,3}.

A prominent cluster of basic residues (R131, R132) delineates the binding site of the substrate C terminus. The strand G918–G916 extends the main-chain interaction sites observed in the TLCK complex structure and was used as primed substrate-docking strand (Fig. 2). There is a prominent hydrophobic S3' pocket formed by W988, Y748, H201, F197, L199 and the aliphatic parts of R131 and R132, and delimited by T209 and Y205 at its floor. The topology of the primed substrate-recognition sites clearly marks tricorn as a carboxypeptidase and explains its preferential di- and tripeptidase activity^{2,3}. The functional role of R131 and R132 in recognizing and fixing the substrate's C terminus is in line with the characteristics of the R131E R132E double mutant. The caseinolytic activity of this mutant is only 10% compared with the wild-type activity, whereas its activity against fluorogenic substrates lacking a negatively

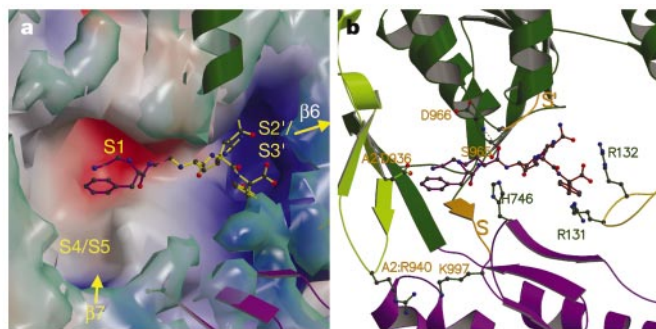


Figure 2 The active site of tricorn protease. **a**, Surface representation of the active site, colour coded according to its electrostatic potential. The 'cut-open' surface (green) indicates an overall acidic S1 site (red), highly basic S2'/S3' substrate-recognition sites (deep blue), and a slightly positive S4/S5 site. **b**, The surface charge distributions related to A2:D936 (provided by the diad-related subunit) and D966 (S1); R131 and R132 (S2'/S3'); and A2/R940 and K992 (S4/S5). The unprimed substrate residues (Phe-Lys) correspond to those experimentally determined from a TLCK complex. The primed residues (Arg-Gln-Tyr-O⁻) were modelled semi-empirically by fitting the solvent electron density in addition to energetic considerations, because solvent molecules are known to mimic substrate-binding sites²⁰.

charged C terminus is slightly increased (J.S.K., M.G., R.H. and H.B., unpublished data). In addition, the resulting electrostatic gradient (Fig. 2a) makes it unlikely for a partly digested substrate to leave the chamber of the active site, thus explaining the processivity of tricorn³.

The comparison with POP, including the open Velcro topology⁵, suggests an important role of the β propellers for substrate access to and product exit from the active site. Both the $\beta 6$ and $\beta 7$ propeller axes are directed towards the active site of the protein. The otherwise direct connection from the chamber of the active site to the exterior through the $\beta 6$ propeller is obstructed by the arginine anchor (R131 and R132), which is located on a helical loop containing three glycines (G126, G130 and G139) lacking any protein contacts (Fig. 2b). In the related *Sulfolobus* tricorn homologue, a segment of seven residues is deleted in this region, thus rendering the $\beta 6$ propeller channel continuously open. Given these observations, we propose that the $\beta 6$ propeller channel represents a gated exit from the catalytic chamber, with the glycines serving as hinge residues. This role is consistent with the point mutant L184C,

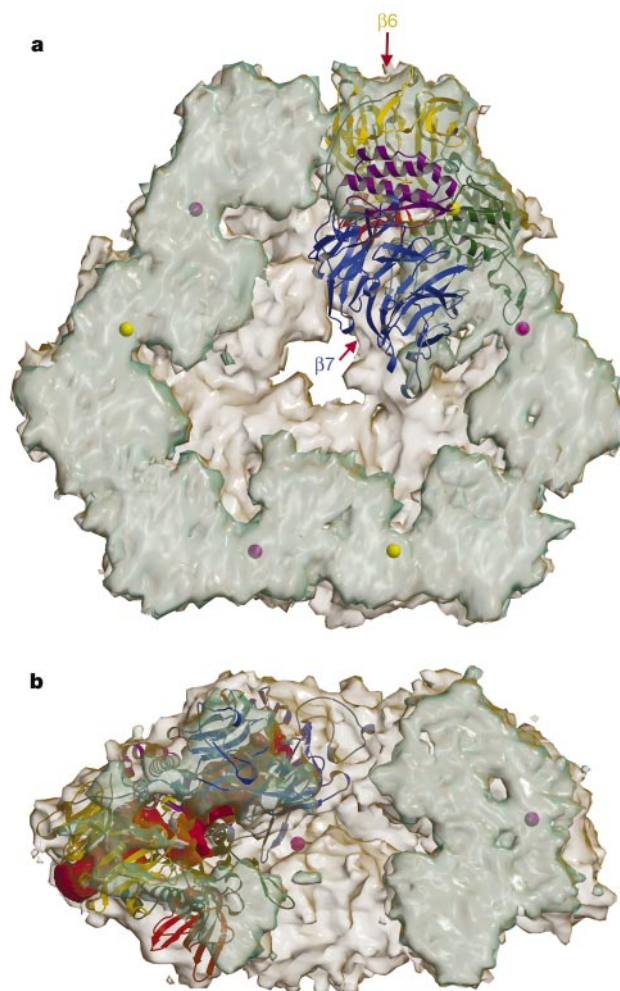


Figure 3 Internal view of tricorn protease. **a**, Cut-open surface representation viewed along the three-fold axis. The central pore, indicated by the grey surface, has a volume of about 97,000 Å³, measuring about 85 and 8 Å in its diameter and height, respectively. It is connected by cavities to the six active sites, indicated as yellow and magenta spheres. **b**, Cut-open surface representation viewed from the side. The $\beta 7$ pseudo-symmetry axes intersect the molecular three-fold (running vertically) at an angle of 57°, thus giving rise to the conical shape of the pore entrance. The red surface indicates internal cavities providing direct access for the substrate through the $\beta 7$ propeller to (blue), and product egress through the $\beta 6$ propeller from (yellow), the active site.

within the $\beta 6$ propeller. The introduced thiol group was modified with maleimide, partially blocking the $\beta 6$ propeller. The activity of this modified mutant enzyme towards fluorogenic substrates is significantly reduced ($<50\%$) compared with the wild-type enzyme (J.S.K., M.G., R.H. and H.B., unpublished data).

The unprimed substrate residues extrapolate towards the $\beta 7$ channel, as described before. Although the channel through the $\beta 7$ propeller is covered by a basic lid at its outside (R369, R414, R645 and K646; Fig. 1b), it provides a significantly shorter route from the outside of the protein to the catalytic chamber ($60 \text{ \AA}$) compared with the alternative route through the central pore ($83 \text{ \AA}$). The latter path to the active site has multiple branches and dead ends. The preferred substrate passage to the active site through the $\beta 7$ channel is consistent with the point mutant R414C, located in the $\beta 7$ channel. Derivatization of this introduced thiol group with maleimide markedly dropped the fluorogenic activity of this mutant to about 50% of the wild-type activity (J.S.K., M.G., R.H. and H.B., unpublished data). A similar gated substrate filter mechanism through a seven-bladed β propeller has been suggested for the prolyl oligopeptidase^{5,9}.

Electron microscope data revealed that 20 tricorn hexamers can assemble as icosahedral capsids *in vivo*^{4,10}. As visual comparison of electron micrographs and the molecular model suggests, the contacts between the hexamers are mediated by three of the six $\beta 6$ propellers. The molecular two-fold axis is not a point-group symmetry element of the icosahedron, and the two trimeric halves become quasi-equivalent in the capsid complex. This break of symmetry translates functionally into a directed substrate flow from the outside of the capsid to the inside; only this direction is consistent with the proposed role of the $\beta 6$ propeller.

Tricorn reportedly cooperates with three additional proteins, termed interacting factors F1, F2 and F3, to degrade oligopeptides sequentially to yield free amino acids³. F1 is a prolyl iminopeptidase with 14% sequence identity to the catalytic domain of POP, which has an additional propeller domain⁵. Guided by the structural scaffold of this latter structure, we speculate that F1 docks onto the six-bladed β propeller of the tricorn core protein. As in POP, substrate enters F1 through the propeller channel in this model. Previously, a physical interaction of F1 with tricorn was reported¹¹. Preliminary data from dynamic light scattering do suggest a low-affinity binding of three F1 proteases with one tricorn hexamer (P. Götting, M.G., J.-S.K. and H.B., unpublished data), consistent with the capsid assembly.

Similarly, there is evidence for functional interaction of tricorn with the proteasome³. If this functional interaction is paralleled by a physical interaction, the exit of product from the proteasome should be aligned with the substrate entrance into the tricorn. This proposed interaction has to be weak, presumably transient (for example, substrate induced), or an auto-inhibitory proteasome-tricorn tunnel would result. Although such a physical interaction would be consistent with the geometric dimensions of both molecules¹², its existence needs to be experimentally confirmed and characterized.

Circumstantial evidence exists for the presence of selected elements of the modular tricorn protease in humans, including their cofactor proteins¹³. Functional tricorn analogues might be difficult to detect in eukaryotes, however. Different from the proteasome, tricorn is built not from a single folding domain, but from five, raising the possibility that tricorn analogues might assemble non-covalently from different gene products, thereby possibly having additional functionalities. Because we propose functional roles for the β propellers in substrate access and product egress, these modules are likely candidates to mediate the interaction with other factors, and might associate as well with these proteins as with the tricorn proteolytic core subdomains. In fact, the eukaryotic D1P of photosystem II is proof of the existence of isolated subdomains (PDZ, C1, C2)⁶. The search for the presence of tricorn

analogues in mammals must continue, perhaps assisted by selective inhibitors designed on the basis of the structural principles presented here. □

Methods

Cloning, protein expression and crystallization

All DNA manipulation was performed according to standard techniques¹⁴. The gene encoding the tricorn proteinase was amplified from *T. acidophilum* genomic DNA by polymerase chain reaction (PCR). The protein was expressed in *Escherichia coli* strain BL21(DE3), and purified by heat treatment followed by ion exchange and size-exclusion chromatography. Crystals were grown by the hanging-drop vapour-diffusion method by mixing equal volumes of concentrated protein (20 mg ml^{-1}) and reservoir solution (25–30% isopropanol 2-morpholinoethanesulphonic acid (MES), pH 6.0) at 20°C . The crystals contain 51% solvent and belong to space group $P2_1$ with unit cell dimensions $a = 96 \text{ \AA}$, $b = 246 \text{ \AA}$, $c = 160 \text{ \AA}$ and $\beta = 105^\circ$.

Structure determination

The structure was determined by a combination of single isomorphous replacement with anomalous scattering (SIRAS), multiwavelength anomalous diffraction (MAD), and NCS averaging, details of which are described elsewhere (see Supplementary Information Table 1). The molecular centre was unambiguously determined by employing the local 3-2 symmetry. This information served to identify a $\text{Ta}_6\text{Br}_{12}$ derivative obeying the 3-2 symmetry. Real-space electron density averaging was performed with MAIN in combination with CCP4 routines^{15,16}. The final back-transformation R -value was 20.8% at $2.0 \text{ \AA}$ resolution. The phases derived from the NCS-averaged electron density were employed to calculate an anomalous difference Fourier map using SeMet-substituted tricorn crystal data, thus allowing straightforward sequence assignment during model building (see Supplementary Information Table 1). The protein model was refined by X-PLOR and CNS¹⁷ with current crystallographic values of $R_{\text{cryst}} = 24.8\%$, $R_{\text{free}} = 25.9\%$, root mean square difference of bonds $\text{r.m.s.d.}_{\text{bond}} = 0.08 \text{ \AA}$ and $\text{r.m.s.d.}_{\text{angle}} = 1.0^\circ$. The current model comprises 49,068 nonhydrogen protein atoms and 2,394 water molecules with average crystallographic temperature factors of $B_{\text{protein}} = 31.1 \text{ \AA}^2$ and $B_{\text{water}} = 33.3 \text{ \AA}^2$, respectively. The average deviation of NCS-related subunits is $0.01 \text{ \AA}$.

Received 8 June; accepted 3 October 2001.

- Wenzel, T., Eckerskorn, C., Lottspeich, F. & Baumeister, W. Existence of a molecular ruler in proteasomes suggested by analysis of degradation products. *FEBS Lett.* **349**, 205–209 (1994).
- Tamura, T. *et al.* Tricorn protease—the core of a modular proteolytic system. *Science* **274**, 1385–1389 (1996).
- Tamura, N., Lottspeich, F., Baumeister, W. & Tamura, T. The role of tricorn protease and its aminopeptidase-interacting factors in cellular protein degradation. *Cell* **95**, 637–648 (1998).
- Walz, J. *et al.* Tricorn protease exists as an icosahedral supermolecule *in vivo*. *Mol. Cell* **1**, 59–65 (1997).
- Fülöp, V., Böcskei, Z. & Polgár, L. Prolyl oligopeptidase: an unusual β -propeller domain regulates proteolysis. *Cell* **94**, 161–170 (1998).
- Liao, D.-I., Qian, J., Chisholm, D. A., Jordan, D. B. & Diner, B. A. Crystal structures of the photosystem II D1 C-terminal processing protease. *Nature Struct. Biol.* **7**, 749–753 (2000).
- Dryjanski, M., Kosley, L. L. & Pietruszko, R. N-tosyl-L-phenylalanine chloromethyl ketone, a serine protease inhibitor, identifies glutamate 398 at the coenzyme-binding site of human aldehyde dehydrogenase—evidence for a second naked anion at the active site. *Biochemistry* **37**, 14151–14156 (1998).
- Ito, N. *et al.* Novel thioether bond revealed by a $1.7 \text{ \AA}$ crystal structure of galactose oxidase. *Nature* **35**, 87–90 (1991).
- Fülöp, V., Szeltnér, Z. & Polgar, L. Catalysis of serine oligopeptidases is controlled by a gating filter mechanism. *EMBO Rep.* **1**, 277–281 (2000).
- Walz, J., Koster, A. J., Tamura, T. & Baumeister, W. Capsids of tricorn protease studied by electron cryomicroscopy. *J. Struct. Biol.* **128**, 65–68 (1999).
- Tamura, T., Tamura, N., Lottspeich, F. & Baumeister, W. Tricorn protease (TRI) interacting factor 1 from *Thermoplasma acidophilum* is a proline iminopeptidase. *FEBS Lett.* **398**, 101–105 (1996).
- Groll, M. *et al.* Structure of the 20S proteasome from yeast at $2.4 \text{ \AA}$ resolution. *Nature* **386**, 463–471 (1997).
- Stolze, L. *et al.* Two new proteases in the MHC class I processing pathway. *Nature Immunol.* **1**, 413–418 (2000).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989).
- Turk, D. in *Chemistry* (Technische Univ. München, Munich, 1992).
- Collaborative Computational Project No. 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Brünger, A. T. *et al.* Crystallography and NMR system—a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
- Kraulis, P. J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Huang, K., Lu, W., Anderson, S., Laskowski, M. J. & James, M. Water molecules participate in proteinase-inhibitor interactions: crystal structures of Leu18, Ala18, and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B. *Protein Sci.* **4**, 1985–1997 (1995).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We acknowledge financial support of the Deutsche Forschungsgemeinschaft and of Roche Diagnostics. P. Götting and R. Ramachandran helped with biochemical analyses. We thank G. Bourenkov and H. Bartunik, and G. Leonard for help with synchrotron data collection at DESY BW6 (Hamburg) and ESRF ID14-4 (Grenoble), respectively.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence should be addressed to H.B. (e-mail: hbs@biochem.mpg.de). The coordinates of the tricorn protease have been deposited in Protein Data Bank under accession code 1K32.

addendum

An efficient room-temperature silicon-based light-emitting diode

Wai Lek Ng, M. A. Lourenço, R. M. Gwilliam, S. Ledain, G. Shao & K. P. Homewood

Nature **410**, 192–194 (2001).

Silicon light-emitting diodes (LED) show light emission at the bandgap energy of silicon with efficiencies approaching those of standard III–V emitters: 0.1% for planar devices (our Letter) and about 1% when total internal reflection is minimized by surface texturing¹. We point out here an additional example of a silicon device also showing light emission at the bandgap². The authors described devices made by the SACMOS-3 process and focus the bulk of the paper on visible emission under reverse bias. However, they also report briefly on a device operated under forward bias giving efficiencies of around 0.01%, although no explanation of the mechanism is given. It is now becoming clear that crystalline silicon, when appropriately engineered, is capable of supporting efficient light emission, opening up many significant applications. □

1. Green, M. A., Shao, J., Wang, A., Reece, P. J. & Gal, M. Efficient silicon light-emitting diodes. *Nature* **412**, 805–808 (2001).
2. Kramer, J. *et al.* Light-emitting devices in industrial CMOS technology. *Sensors Actuators A37–A38*, 527–533 (1993).

erratum

Warm tropical sea surface temperatures in the Late Cretaceous and Eocene epochs

Paul N. Pearson, Peter W. Ditchfield, Joyce Singano, Katherine G. Harcourt-Brown, Christopher J. Nicholas, Richard K. Olsson, Nicholas J. Shackleton & Mike A. Hall

Nature **413**, 481–487 (2001).

In this Article, the temperature scale in Figure 3i should have been the same as in Figure 3g. □

corrections

Self-assembled monolayer organic field-effect transistors

Jan Hendrik Schön, Hong Meng & Zhenan Bao

Nature **413**, 713–716 (2001).

The values of the transconductance in Table 1 and in the text (page 715, second paragraph) are incorrect. The values should be divided by ten. The data plotted in Figs 2 and 3 are correct and the conclusions are not affected. □

Ordered nanoporous arrays of carbon supporting high dispersions of platinum nanoparticles

Sang Hoon Joo, Seong Jae Choi, Ilwhan Oh, Juhyoung Kwak, Zheng Liu, Osamu Terasaki & Ryong Ryoo

Nature **412**, 169–172 (2001).

We inadvertently omitted to cite an earlier reference alongside ref. 8 (G. Che, B. Lakshmi, E. R. Fisher and C. R. Martin *Nature* **393**, 346–349; 1998), which was published in 1995 (and not 2000 as printed). Also, our suggestion that using the pores in a microporous material as templates could be a way in which to produce nanoscale materials has been discussed before (see, for example, C. R. Martin *Science* **266**, 1961–1966 (1994) and J. C. Hulthén & C. R. Martin *J. Mater. Chem.* **7**, 1075–1087 (1997)). □

The timing of the last deglaciation in North Atlantic climate records

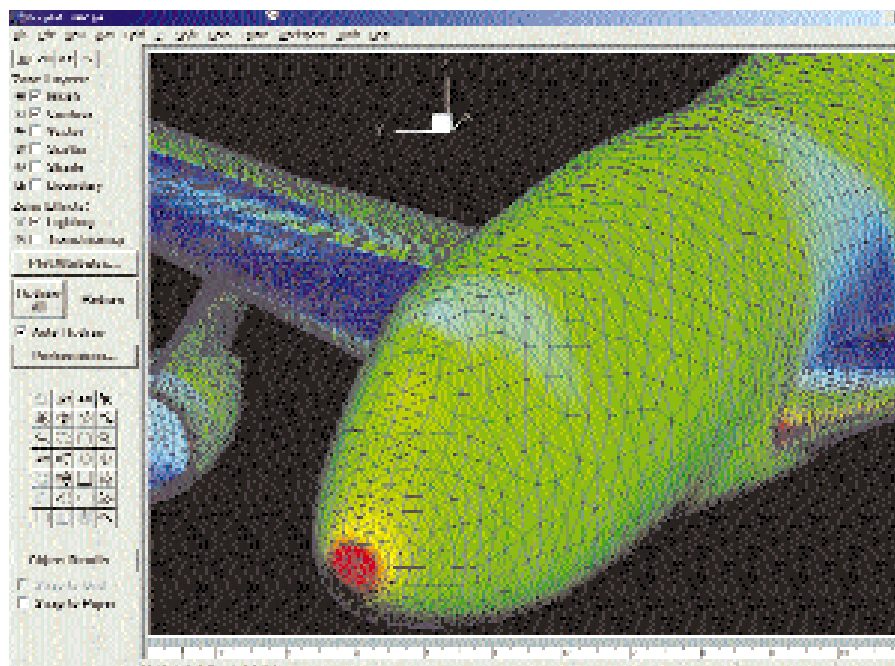
Claire Waelbroeck, Jean-Claude Duplessy, Elisabeth Michel, Laurent Labeyrie, Didier Paillard & Josette Duprat

Nature **412**, 724–727 (2001).

We directly used the observed leads of sea surface temperature with respect to air temperature (dated in calendar years), whereas the air temperature calendar ages should have been converted into ¹⁴C ages, with reservoir ages computed as the difference between marine and atmospheric ¹⁴C ages. Taking this into consideration, apparent surface-water ages are 1,180 ± 630 to 1,880 ± 750 years at the end of the Heinrich 1 surge event (14,500 years BP) and 930 ± 250 to 1,050 ± 230 years at the end of the Younger Dryas cold episode. This does not change the discussion and conclusions. □

Looking for soft options?

A baker's dozen of the latest software releases for the lab.



Don't lose the plot: a free demo of Tecplot 9.0 can be downloaded from polyhedron.com.

Tecplot 9.0

Polyhedron Software www.polyhedron.com
Travel in the third dimension

Tecplot 9.0 runs on graphics cards of the type normally used in computer games and can be used to achieve high-end three-dimensional data visualization on Unix workstations and PCs. The Tecplot program displays the results of finite element or other simulations using the same grid as the simulator, in contrast to programs that interpolate the results onto a cartesian grid, a method that gains speed but loses detail. Polyhedron claims that Tecplot's improved speed allows users to 'fly' interactively through large two- and three-dimensional data sets consisting of millions of points, while fully retaining the integrity of their data. The program is available for Windows 95, 98, NT, 2000, Linux and Unix.

OptiQuant 4

Packard Bioscience
www.packardbioscienc.com
Operating system for Cyclones

OptiQuant is the operating and analysis software that runs the Cyclone storage phosphor system and performs analysis of blots, gels, tissue sections and TLC plate images. Version 4 is compatible with Windows 98, NT and Windows 2000 operating systems. With the purchase of a Cyclone, OptiQuant can be loaded on to all other workstations in a user's lab and

office and used to analyse and publish data and images. A Cyclone GeneArray system is also available, integrating OptiQuant version 4 and QuantArray 3.0.

Pathways 4

ResGen www.resgen.com
Java-based microarray analysis

Pathways 4 universal microarray analysis software provides a comprehensive set of tools for the study of differential gene expression. The Java-based program allows cross-platform use and integrates with plug-ins. A brochure outlining how Pathways 4 imports microarray data from images, spreadsheets, GEML files and other sources and describing analysis and filtering options for data sets is available via pathways.resgen.com.

ComIML

Windmill Software www.windmill.co.uk
Driving ambition

This instrument driver handles most equipment that can be connected to a PC's COM port. It is bundled with Windmill's data logging, charting and control applications. Users can send data to other Windows programs such as Excel spreadsheets or LIMS applications. ComIML runs under Windows 95, 98, NT, 2000 or ME and can send either ASCII or binary data over many serial instruments, including gas analysers, pH transmitters, titra-

tors, particle analysers, conductivity meters, laboratory scales, GPS and sonar. The set-up program can be downloaded from the company's website to enable users to enter their instrument settings and ensure that they can receive data before purchasing the driver.

AQM 2001

Kinetic Imaging www.kineticimaging.com
Versatile multiparameter imaging software

Designed to provide data-rich imaging of living cells, the AQM 2001 program offers enhanced tools for multichannel, multiwell plate scanning. Advanced data storage, retrieval, analysis and presentation tools such as an AVI movie maker are supplied as standard. Applications include multichannel imaging with GFP and its derivatives. Optional modules for tracking movement and spot analysis are available. The company can also supply complete live cell imaging systems and supports all leading cameras, microscopes and accessories.

CISPro 2000

ChemSW www.chemsw.com
Search by structure

The latest version of ChemSW's CIS (Chemical Inventory System) software includes a new structure search capability. Databases can now be searched by structure and substructure, using the .mol file format. Structures can be pasted directly into CISPro search windows from a structure database or draw program. The structure search feature is a standard part of CISPro Version 5 and is also available as an upgrade to users of earlier versions. Client/server and web versions for both Oracle and MS SQL Server applications are available.

SNP Primer Design

Pyrosequencing www.pyrosequencing.com
Optimize SNP analysis

This web-based software assay design tool for use in DNA analysis will evaluate and select primers tailored for use with the company's Pyrosequencing PSQ 96 and PTP systems. Researchers performing single nucleotide polymorphism analyses can use SNP Primer Design to optimize the selection of individual sequencing primers. The software designs primers in either forward, reverse or both directions, computes the oligonucleotide's melting temperature, predicts secondary structures and potential primer interactions and ranks primers according to their suitability for each SNP assay. Analyses can be stored in HTML format on local computers.

Labworks

PerkinElmer Instruments

www.perkinelmer.com

Open learning

The Labworks laboratory information management system user website contains application-specific user training videos, online documentation, a 'frequently asked questions' section and message boards. It also functions as a virtual university classroom. Users can review a course calendar, choose topics that interest them and register online. They can then log on to the user site at the scheduled time and receive training on a specific Labworks topic. The message board connects users to PerkinElmer technicians and to each other, and contains helpful suggestions and implementation techniques.

web Mathematica

Wolfram Research

www.wolfram.com

Interactive solutions

This software release takes the computational power of Mathematica on to the web. It can be used to build custom websites that provide specialized calculations and to compute, visualize and analyse data using a web browser. Courseware and interactive books and technical documentation can also be published for web access. The program is compatible with most web servers and can be used with Windows 95, 98, ME, NT, 2000 and Intel-based Linux platforms. Two versions are available: the 'professional' is designed for development of sophisticated Intranet and Internet sites, and 'amateur' is for individuals who wish to showcase their work over the web.

Gene Construction/ Gene Inspector

Textco

www.textco.com

Enhancements and enticements

Users of the Gene Construction Kit and Gene Inspector can download free maintenance updates and database research tables from Textco's redesigned website. Restriction enzyme and Prosite databases are custom-formatted and also available free of charge to cus-

tomers. Prospective customers are offered free demonstration versions and tutorials. The Gene Construction Kit is used for DNA manipulation, cloning design and creating illustrations. Gene Inspector is an electronic lab notebook and sequence analysis product.

Design-Ease 6.0

Stat-Ease

www.statease.com

Experiment with entry-level software

This dedicated package of experimental design tools allows inexperienced experimenters to improve their product or process by revealing the vital factors in the processes and the interactions between them. It offers an intuitive interface, an expanded help system, a wide variety of designs with flexibility for modification, graphics and evaluation capabilities such as annotated ANOVA reports. The package features standard two-level and fractional factorials, general (multilevel) factorials, high-resolution irregular fractions, Taguchi orthogonal arrays and Plackett-Burman designs. A complete array of diagnostic graphs is displayed, enabling users to check statistical assumptions and detect possible outliers in the data. A new Box-Cox diagnostic plot provides advice on the best response transformation.

SDMS 5.0

NuGenesis

www.nugenesis.com

Collective effort

SDMS (Scientific Data Management System) is a suite of software for the collection, storage and management of laboratory data. Data from scientific instruments are captured and can be catalogued according to protocols and projects. The program will also automatically aggregate information from disparate sources. Version 5.0 features web technology so that projects and reports can be saved as hyperlinks in a favourites folder in either Netscape or Internet Explorer. Projects and NuGenesis databases worldwide can be searched using standard browsers. Dynamically linked raw data, report data and interpreted data can be shared with colleagues via the Internet or sent



Easy does it: designing starts from first principles.

to a word processing, spreadsheet or presentation application. Users wishing to locate a report manually can hyperlink back to NuGenesis SDMS from a compound document and print key traceable information. Annotations or comments can be added to reports and are tracked, with an annotation history maintained in the database and in an audit trail.

HyperRF

Shimadzu

www.shimadzu.com

Spectral display

Compatible with Windows 95, 98 and NT, this software package for Shimadzu's RF-5301PC and RF1501 spectrofluorophotometers enables multi-component sample analysis, creates three-dimensional spectral plots and comes with intracellular Ca^{2+} determination capabilities. It offers full NT security as well as GLP/GMP compliance functions, including password protection and report trails. Data management features include data sharing, import and export, and customized report layouts.

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

Proteasome inhibition?

- ~ epoxomicin ~
- ~ PR39 ~ lactacystin ~
- ~ clasto-lactacystin ~ β ~
- ~ lactone ~ P131 ~
- ~ MG-132 ~ MG-262 ~

www.proteasome.com

AFFINITY Research Products Ltd.
Mamhead Castle, Exeter, EX6 8HD, U.K.
T.+44(0) 1626 891010 - F+44(0) 1626 891090
- Email info@proteasome.com -

**DIVERSE
SMALL
MOLECULES
READY FOR
SCREENING**



**10K-50K
Compounds
in Plates**

- Next Day Delivery*
- Competitively Priced

(Some Restrictions Apply)

Chembridge Corporation • 800.964.6143
858.451.7400 • 858.451.7401 FAX • E-mail: chem@chembridge.com

www.chembridge.com

Polyclonal Antibodies & Custom Peptides

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

ResGen™
Invitrogen Corporation

U.S. or Canada 800-533-4363
FAX 256-536-9016

Check out our web site at
www.resgen.com

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:
Leonie Welss (4954)
European Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)
Holland/ Italy: Nevin Bayoumi (4978)
Scandinavia: Silke Opstrup (4994)
Spain/ Portugal: Leonie Welss
(4954)

Production Manager:

Billie Franklin
To send films and materials use
London address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakataamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejp.com
Japan Manager: Kate Cowan

naturejobs

Demographic shifts

One of the forces that could have a major impact on scientific employment in the near future is also one of the most obvious: retirement. There are signs that a disproportionate number of life-science faculty members will leave their posts in the next few years, making room for younger replacements.

For example, about a third of the academic staff in the University of Cambridge's pharmacology department is over 60. One professor will retire this year, three in 2003, and another in 2006. These departures will mean new recruits — assuming that the department maintains its size.

In *Naturejobs* earlier this year (see pages 4–5; 19 July 2001), several European immunologists anticipated a similar trend in their field. And in this issue, some Italian scientists predict an imminent shortfall in the young blood needed to replace those who will retire over the next decade (see overleaf). Italy's academic system appears to be exacerbating the problem — many young Italian scientists are finding better funding and career prospects abroad.

So far, the overall evidence is anecdotal and scattered. But as it accumulates, it grows more convincing. The question is, what to do about the expected deficit. It might be irresponsible to ask universities to start churning out huge numbers of PhDs — that would only recreate the clogged pipeline that newly minted life scientists have travelled through in recent years.

But it may be appropriate to encourage young scientists not to abandon all hope of ever attaining an academic career. And for those countries experiencing a scientific brain drain, it would be prudent for them to acknowledge the problem before the glut of scientific talent turns into a shortage.

Paul Smaglik

Naturejobs editor



Contents

SPECIAL REPORT

Europe faces up to a
brain drain

p4

WWW.NATUREJOBS.COM

Career centre

Information on the scientific
job market

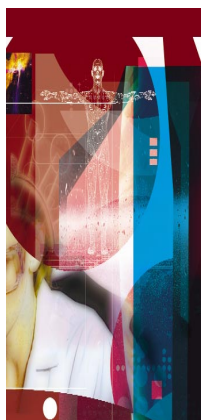
FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



The exodus of scientists from Italy reflects problems several European countries are trying to tackle. Are their efforts too little, too late? asks Alexander Hellemans.

Discouraged by his efforts to find a job in his native Italy, astronomer Marco Bruni left the country in 1992 and took a job in London. He has remained abroad ever since, and now works at Portsmouth University in England. He feels much more secure about his scientific career since he left and doesn't anticipate returning. "There is nothing like the Italian situation here that forces people to go abroad," Bruni says. "In Italy, many people can't find a job at all."

The scientific brain drain is by no means limited to Italy. It's a problem faced to varying degrees by other European countries including Spain, France and Germany. Each country is taking its own approach to turning the tide, but Italy's situation encapsulates the problems that other states are hoping to escape. The shortage of secure jobs isn't caused by crowds of researchers competing for a place — far from it.

According to Sveva Avveduto, who researches science and education policy at Italy's National Research Council (CNR), the population of active scientists in Italy is now quite old, with about 30% expected to retire by 2005.

To keep numbers steady, Italy's universities should be producing 12,000 new researchers every year instead of the current 4,000. On top of this, several hundred scientists leave the country each year, says Avveduto, and the chances that they will return are slim.

ITALIAN DEFICIT

Many factors contribute to a scientist's decision to leave Italy, including the relatively low level of investment in science, a cumbersome bureaucracy and an advancement system that many say has more to do with cronyism than quality (see *Nature* **412**, 264–265; 2001).

These factors often lead to low pay and an uncertain future, especially for young scientists. Even 10 years into their career, some Italian researchers have to get by on 1,000 euros (US\$886) a month, says astronomer Sandra Savaglio of Johns Hopkins University in Baltimore, Maryland. Bruni adds that the pay differential in Italy is wider than in other European states, with young researchers earning as little as a quarter of a more senior staff member's salary.

The problem is so profound that even scientists who land secure jobs sometimes leave Italy for better opportunities. Savaglio, for example, recently accepted

her current position at Johns Hopkins after a year in a permanent position at the Rome Observatory.

The differences become more noticeable when Italian scientists spend time in new posts abroad. Higher salaries aren't the only incentive to move, says Bruni. "I found myself treated as a full staff member, and not like an aged student," he says.

Mara Lorenzi, an endocrinologist at Harvard Medical School who left Italy to do medical research in the United States some 25 years ago, says that Italy did not seem to keep pace with the revolution in the life sciences that happened in the past 50 years. "The universities did not actively create opportunities for research," she remembers. Italian biomedical researchers still head for the United States, she says, although no longer in such great numbers.

CULTURAL DIVIDE

Perhaps most galling to young Italian researchers is what many describe as an anachronistic system of distributing jobs in research. It's who you know, not what you know, that counts, say several who have left the country. Applying for research positions abroad comes like a breath of fresh air.

"The most important thing here is that you are considered a good researcher," Bruni says of Britain. If you want to find a position in Italy, he says, you have to take a different approach and adapt to the hierarchical structure. That often means garnering support from a single research director who can make or break a younger scientist's career.

Italy's problem with patronage is a warning to other countries where a similar culture contributes to the brain drain, especially Spain. "The typical way to make your career is to be next to a professor who is active," says Enric Banda, Spain's former secretary of state for universities and research, now secretary general of the European Science Foundation (ESF). "You are in his or her shadow, and if you wait long enough, eventually you will get a position. As we say in Spain, you have to 'keep your seat warmed'."

Many are not prepared to wait. Italians form the largest group of foreign scientists at the Space Telescope Science Institute, the Baltimore-based centre of operations for the Hubble Telescope, says Massimo Stiavelli, who joined the institute six years ago. Large Italian research communities also exist at the European

Sandra Savaglio left a permanent position in Rome for better opportunities in the United States.





Cash for southern comfort

Southern Observatory in Garching, Germany, and CERN, the European Laboratory for Particle Physics near Geneva. Some 200 Italian researchers work in the United States for the National Institutes of Health.

Although many young European scientists spend some time abroad as part of their career, in most places the numbers arriving and leaving stay roughly even. Sami Mahroum, an associate of the Dublin-based CIRCA Group Europe, a research and technology policy consultancy that has studied Europe's brain drain, says that the United Kingdom and Ireland, for example, participate more actively in the international circulation of scientists and academics than Italy.

Overall, though, Europe is less attractive for scientists than the United States, says Banda. "One of the main reasons is flexibility, which we don't have," he says. When he was a hands-on scientist in Spain, he sent a few of his students to the United States as postdocs. "Because they were good, they never came back."

The problem isn't only in southern Europe: even traditionally robust countries such as Germany are feeling the pinch. Wilhelm Krull, secretary general of the Hannover-based Volkswagen Foundation, estimates that about 60% of German scientists doing postdocs abroad don't return. Meanwhile, he says, a sharp decline in the number of students taking up science compounded with a brain drain is already causing a shortage of scientists.

EFFORTS TO REVERSE THE TIDE

Individual countries are making their own attempts to reverse the tide. Some, such as France, are increasing research budgets. Spain launched the Ramón y Cajal programme earlier this year, hoping to lure scientists back from abroad by offering five-year contracts with attractive salaries and the chance of managing their own projects (see *Nature* **413**, 556; 2001). Italy is setting up a similar programme offering four-year contracts, reports Avveduto. In addition, private foundations, such as the Wellcome Trust in Britain and the Volkswagen Foundation, can fund young scientists, as Krull points out, in competition with the financially powerful US foundations.

But more reforms will almost certainly be necessary to reverse Italy's brain drain. Recent developments, such as the planned reduction of science funding, which at 0.6% of GNP of public funding is among the

Although public funding of science in Italy has been largely stagnant, the government is pumping lire into targeted regions. Naples, with its history of hosting genetics research, is gaining funds for post-genomic research. Recently, the government granted 40 billion lire (\$18.2 million) to the research consortium BioGeM, of which half is tagged for a new institute to be built south of Avellino. The institute, scheduled for completion in 2004, will house up to 200 researchers.

This funding has helped Emilia De Santis leave her veterinary practice and join a team at BioGeM working on mice to investigate the genetic component of human thyroid disorders. "For the south

of Italy, this is a new possibility," she says.

But because of the way public funding is distributed — money is allocated for different programmes every year, without continuity for a specific programme — ensuring continued support for BioGeM will be problematic, says BioGeM president Roberto Di Lauro. "What Italy needs is a secure source of money which is reliable," he says.

Private initiatives, such as Italy's Telethon Foundation and the Italian Association for Cancer Research are both providing such a source. "Perhaps the best Italian research is done on cancer and on genetic diseases because there are these two foundations," Di Lauro suggests.

A.H.

◆ biogem.iigb.na.cnr.it

Mara Lorenzi believes that Italy failed to keep pace with scientific developments, which means that other European institutes, such as CERN (left), prove an attractive draw.

lowest in Europe, will not help matters. And the recent turmoil surrounding the nomination of directors for new institutes under the CNR shows that Italy is still slow to lose its old habits (see *Nature* **414**, 133; 2001) — the first 22 selected had all been CNR directors, and well-qualified applicants were passed over for reasons that were not made clear.

As a step in the right direction, a law allowing the appointment of non-Italians for positions with the CNR or in public research was passed last year — appointment of non-nationals has long been routine in other countries, for example with the French national research agency CNRS or Germany's Max Planck Institutes. For Avveduto, the influx of non-Italian scientists would be a healthy 'brain gain'.

Alexander Hellemans is a science writer based in Naples.



Emilia De Santis sees hope for Italy's south.